DIFFERENTIAL FORMS AND THE GEOMETRY OF GENERAL RELATIVITY

DIFFERENTIAL FORMS AND THE GEOMETRY OF GENERAL RELATIVITY

TEVIAN DRAY
OREGON STATE UNIVERSITY
CORVALLIS, USA

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
AN A K PETERS BOOK

The following figures appear in The Geometry of Special Relativity [2] and are reproduced with permission: Figures 1.5, 7.2(a)-(b), 7.3, and 18.13(a)-(b).

The following figures are from the online text The Geometry of Vector Calculus [1] and are reproduced with permission under the Creative Commons by-nc-nd license: Figures 1.1, 1.2, 2.1, 2.2, 12.1, 12.2, 16.1, 16.3, 16.4, 13.7, 17.2, 18.1, 18.2, 18.3, and 18.4.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140902

International Standard Book Number-13: 978-1-4665-1032-6 (eBook - PDF)

Differential geometry is just advanced vector calculus.

Curvature = Matter.

Contents

List of Figures and Tables

Preface

This book contains two intertwined but distinct halves, each of which can in principle be read separately. The first half provides an introduction to general relativity, intended for advanced undergraduates or beginning graduate students in either mathematics or physics. The goal is to describe some of the surprising implications of relativity without introducing more formalism than necessary. "Necessary" is of course in the eye of the beholder, and this book takes a nonstandard path, using differential forms rather than tensor calculus, and trying to minimize the use of "index gymnastics" as much as possible.[1] This half of the book is itself divided into two parts, the first of which discusses the geometry of black holes, using little more than basic calculus. The second part, covering Einstein's equation and cosmological models, begins with an informal crash course in the use of differential forms, and relegates several messy computations to an appendix.

The second half of the book (Part III) takes a more detailed look at the mathematics of differential forms. Yes, it provides the theory behind the mathematics used in the first half of the book, but it does so by emphasizing conceptual understanding rather than formal proofs. The goal of this half of the book is to provide a language to describe curvature, the key geometric idea in general relativity.

[1] For the expert, the only rank-2 tensor objects that appear in the book are the metric tensor, the energy-momentum tensor, and the Einstein tensor, all of which are instead described as vector-valued 1-forms; the Ricci tensor is only mentioned to permit comparison with more traditional approaches.

Parts I and II: General Relativity

As with most of my colleagues in relativity, I learned the necessary differential geometry the way mathematicians teach it, in a coordinate basis. It was not until years later, when trying to solve two challenging problems (determining when two given metrics are equivalent, and studying changes of signature) that I became convinced of the advantages of working in an orthonormal basis. This epiphany has since influenced my teaching at all levels, from vector calculus to differential geometry to relativity. The use of orthonormal bases is routine in physics, and was at one time the standard approach to the study of surfaces in three dimensions. Yet no modern text on general relativity makes fundamental use of orthonormal bases; at best, they calculate in a coordinate basis, then reinterpret the results using a more physical, orthonormal basis.

This book attempts to fill that gap.

The standard basis vectors used by mathematicians in vector analysis possess several useful properties: They point in the direction in which the (standard, rectangular) coordinates increase, they are orthonormal, and they are the same at every point. No other basis has all of these properties; whether working in curvilinear coordinates in ordinary, Euclidean geometry, or on the curved, Lorentzian manifolds of general relativity, some of these properties must be sacrificed.

The traditional approach to differential geometry, and as a consequence to general relativity, is to abandon orthonormality. In this approach, one uses a *coordinate basis*, in which, say, the basis vector in the θ direction corresponds to the differential operator that takes θ-derivatives. In other words, one defines the basis vector $\vec{\boldsymbol{e}}_\theta$ by an equation of the form

$$\vec{\boldsymbol{e}}_\theta \cdot \vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial \theta}.$$

Physics, however, is concerned with measurement, and the physically relevant components of vector (and tensor) quantities are those with respect to an *orthonormal basis*. The fact that angular velocity is singular along the axis of symmetry is a statement about the use of angles to measure "distance", rather than an indication of a physical singularity. In relativity, where we don't always have a reliable intuition to fall back on, this distinction is especially important. We therefore work almost exclusively with orthonormal bases. Physics students will find our use of normalized vector fields such as

$$\hat{\boldsymbol{\theta}} = \frac{\vec{\boldsymbol{e}}_\theta}{|\vec{\boldsymbol{e}}_\theta|} = \frac{1}{r}\,\vec{\boldsymbol{e}}_\theta$$

familiar; mathematics students probably won't.

In both approaches, however, one must abandon the constancy of the basis vectors. Understanding how the basis vectors change from point to point leads to the introduction of a *connection*, and ultimately to *curvature*. These topics are summarized informally in Section 6.1, with a detailed discussion deferred until Part III.

We also follow an "examples first" approach, beginning with an analysis of the Schwarzschild geometry based on geodesics and symmetry, and only later discuss Einstein's equation. This allows the reader an opportunity to master the geometric reasoning essential to relativity before being asked to follow the more sophisticated arguments leading to Einstein's equation. Along the way, we discuss the standard applications of general relativity, including black holes and cosmological models.

No prior knowledge of physics is assumed in this book, although the reader will benefit from familiarity with Newtonian mechanics and with special relativity. This book does however assume a willingness to work with differential forms, which in turn requires familiarity with vector calculus and linear algebra. For the reader in a hurry, the essentials of both special relativity and differential forms are reviewed in Chapters 1 and 6, respectively.

PART III: DIFFERENTIAL FORMS

I took my first course in differential geometry as a graduate student. I got an A, but I didn't learn much. Many of my colleagues, including several non-mathematicians with a desire to learn the subject, have reported similar experiences.

Why should this be the case? I believe there are two reasons. First, differential geometry—like calculus—tends to be taught as a branch of analysis, not geometry. Everything is a map between suitable spaces: Curves and surfaces are parametrized; manifolds are covered with coordinate charts; tensors act on vectors; and so on. This approach may be good mathematics, but it is not very enlightening for beginners. Second, too much attention is given to setting up a general formalism, the tensor calculus. Differential geometry has been jokingly described as the study of those objects which are invariant under changes in notation, but this description is a shockingly accurate summary of the frustrations numerous students experience when trying to master the material.

This part of the book represents my attempt to do something different. The goal is to learn just enough differential geometry to be able to learn

the basics of general relativity. Furthermore, the book is aimed not only at graduate students, but also at advanced undergraduates, not only in mathematics, but also in physics.

These goals lead to several key choices. We work with differential forms, not tensors, which are mentioned only in passing. We work almost exclusively in an orthonormal basis, both because it simplifies computations and because it avoids mistaking coordinate singularities for physical ones. And we are quite casual about concepts such as coordinate charts, topological constraints, and differentiability. Instead, we simply assume that our various objects are sufficiently well-behaved to permit the desired operations. The details can, and in my opinion should, come later.

This framework nonetheless allows us to recover many standard, beautiful results in $\mathbb{R}^3$. We derive formulas for the Laplacian in orthogonal coordinates. We discuss—but do not prove—Stokes' Theorem. We derive both Gauss's *Theorema Egregium* about intrinsic curvature and the Gauss–Bonnet Theorem relating geometry to topology. But we also go well beyond $\mathbb{R}^3$. We discuss the Cartan structure equations and the existence of a unique Levi-Civita connection. And we are especially careful *not* to restrict ourselves to Euclidean signature, using Minkowski space as a key example.

Yes, there is still much formalism to master. Furthermore, this classical approach is no longer standard—and certainly not as an introduction to relativity. I hope to have presented a coherent path to relativity for the interested reader, with some interesting stops along the way.

WEBSITE

A companion website for the book is available at

http://physics.oregonstate.edu/coursewikis/DFGGR/bookinfo

Acknowledgments

First and foremost, I thank my wife and colleague, Corinne Manogue, for discussions and encouragement over many years. Her struggles with the traditional language of differential geometry, combined with her insight into how undergraduate physics majors learn—or don't learn—vector calculus have had a major influence on my increased use of differential forms and orthonormal bases in the classroom.

The use of differential forms, and especially of orthonormal bases, as presented in this book, represents a radical change in my own thinking. The relativity community consists primarily of physicists, yet they mostly learned differential geometry as I did, from mathematicians, in a coordinate basis. This gap is reminiscent of the one between the vector calculus taught by mathematicians, exclusively in rectangular coordinates, and the vector calculus used by physicists, mostly in curvilinear coordinates, and most definitely using orthonormal bases.

I have had the pleasure of working with Corinne for more than a decade to try to bridge this latter gap between mathematics and physics. Our joint efforts to make $d\vec{r}$ the key concept in vector calculus also led to my redesigning my differential geometry and relativity courses around the same idea.

My debt to Corinne is beyond words. She opened my eyes to the narrowness of my own vision of vector calculus, and, as a result, of differential geometry. Like any convert, I have perhaps become an extremist, for which only I am to blame. But the original push came from Corinne, to whom I am forever grateful.

I thank my department for encouraging the development of an undergraduate mathematics course in general relativity, then supporting this course over many years. I am grateful for the support and interest of numerous students, and for their patience as I experimented with several textbooks, including my own.

I am also grateful for the extensive support provided by the National Science Foundation for our work in vector calculus. Although this book is not directly related to those projects, there is no question that it was greatly influenced by my NSF-supported work. The interested reader is encouraged to browse the project websites for the Paradigms in Physics Project (http://physics.oregonstate.edu/portfolioswiki) and the Vector Calculus Bridge Project (http://www.math.oregonstate.edu/bridge), as well as our online vector calculus text [1].

How to Read This Book

There are several paths through this book, with different levels of mathematical sophistication. The two halves, on general relativity (Parts I and II) and differential forms (Part III), can be read independently, and in either order. I regularly teach a 10-week course on differential forms using Part III, followed by a 10-week course on general relativity, using Parts I and II but skipping most of Appendix A (and all of Appendix B). However, there are always a few students who take only the second course, and who make do with the crash course in Section 6.1.

Basic

The geometry of the spacetimes discussed in this book can be understood as geometric models without knowing anything about Einstein's field equation. This path requires only elementary manipulations starting from the line element, together with a single symmetry principle, but does not require any further knowledge of differential forms.

With these basic tools, a detailed study of the Schwarzschild geometry is possible, including its black hole properties, as is the study of simple cosmological models. However, the fact that these solutions solve Einstein's equation must be taken on faith, and the relationship between curvature, gravity, and tidal forces omitted.

Read:

- Chapters 1–5;
- Section 6.3;
- Chapters 9 and 10.

STANDARD

This path represents the primary route through the first half of the book, covering all of the content, but leaving out some of the details. Familiarity with differential forms is assumed, up to the level of being able to compute connection and curvature forms, at least in principle. However, familiarity with (other) tensors is not necessary, provided the reader is willing to treat the metric and Killing's equation informally, as simple products of infinitesimals.

Some further advanced mathematical topics can be safely skipped on this path, such as the discussion of the divergence of the metric and Einstein tensors in Appendix A. The reader who chooses this path may also choose to omit some computational details, such as the calculations of curvature given in Appendix A; such computations can also easily be done using computer algebra systems.

Read:

- Chapters 1–10.

EXPERT

Advanced readers will want to work through most of the computations in Appendix A.

Read:

- Chapters 1–10;
- Appendix A.

A TASTE OF DIFFERENTIAL FORMS

This path represents a quick introduction to differential forms, without many details.

Read one or both, in either order:

- Chapters 11 and 12;
- Sections 6.1 and 6.2.

A COURSE IN DIFFERENTIAL FORMS

This path represents a reasonable if nonstandard option for an undergraduate course in differential geometry. Reasonable, because it includes both Gauss's *Theorema Egregium* about intrinsic curvature and the Gauss–Bonnet Theorem relating geometry to topology. Nonstandard, because it does not spend as much time on curves and surfaces in $\mathbb{R}^3$ as is typical. Advantages to this path are a close relationship to the language of vector calculus, and an introduction to geometry in higher dimensions and with non-Euclidean signature.

Read:

- Chapters 11–20.

Part I

Spacetime Geometry

≻ Chapter 1 ≺

Spacetime

1.1 Line Elements

The fundamental notion in geometry is distance. One can study shapes without worrying about size or scale, but that is topology; in geometry, size matters.

So how do you measure distance? With a ruler. But how do you calibrate the ruler?

In Euclidean geometry, these questions are answered by the Pythagorean Theorem. In infinitesimal form,[1] and in standard rectangular coordinates, the Pythagorean Theorem tells us that

$$ds^2 = dx^2 + dy^2 \tag{1.1}$$

as shown in Figure 1.1. This version of the Pythagorean Theorem not only tells us about right triangles, but also how to measure length along any curve: Just integrate (the square root of) this expression to determine the arclength.

More generally, an expression such as (1.1) is called a *line element*, and is also often referred to as the *metric tensor*, or simply as the *metric*.

The line element (1.1) describes a Euclidean plane, which is flat. What is the line element for a sphere? That's easy: Draw a picture! Figure 1.2 shows two "nearby" points on the sphere, separated by a distance ds, broken up into pieces of lengths $r\,d\theta$ at constant longitude and $r\sin\theta\,d\phi$ at constant latitude.[2] Alternatively, write down the line element for three-dimensional Euclidean space in rectangular coordinates, convert everything to spherical coordinates, then hold r constant. Yes, the algebra is messy.

---

[1] We work throughout with infinitesimal quantities such as dx, which can be thought of informally as "very small," or more formally as differential forms, a special kind of tensor. Some further discussion can be found in Section 11.1.

[2] We use "physics" conventions, with θ measuring colatitude and ϕ measuring longitude, and we measure both "sides" of our right "triangle" starting at the same point, as is appropriate for infinitesimal objects in curvilinear coordinates. For further details, see our online vector calculus text [1].

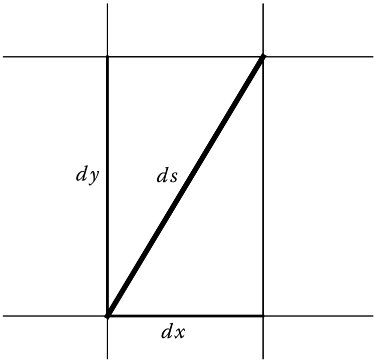


FIGURE 1.1. The infinitesimal version of the Pythagorean Theorem on a plane in rectangular coordinates, illustrating the identity $ds^2 = dx^2 + dy^2$.

The resulting line element is

$$ds^2 = r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) . \tag{1.2}$$

The study of such curved objects is called *Riemannian geometry.*

Both line elements above, (1.1) and (1.2), are *positive definite*, but we can also study geometries which do not have this property. The simplest example is

$$ds^2 = dx^2 - dt^2, \tag{1.3}$$

which turns out to describe special relativity; this geometry turns out to be

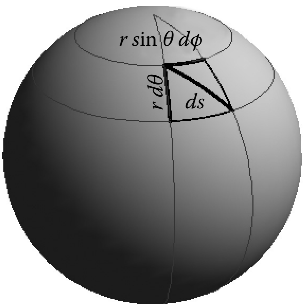


FIGURE 1.2. The infinitesimal version of the Pythagorean Theorem on a sphere, illustrating the identity $ds^2 = r^2(d\theta^2 + \sin^2\theta \, d\phi^2)$.

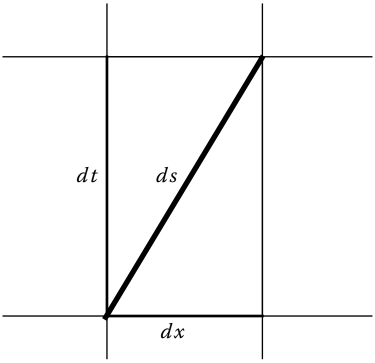


FIGURE 1.3. The infinitesimal version of the Pythagorean Theorem in special relativity.

flat, and is called *Minkowskian*. We classify line elements by their *signature* s, which counts the number of minus signs; $s = 1$ for both special and general relativity. Figure 1.3 shows the infinitesimal Pythagorean Theorem for Minkowski space (special relativity). Yes, this figure looks the same as Figure 1.1, but this geometry is not that of an ordinary piece of paper. We use this representation because Minkowski space turns out to be flat, but the minus sign in the Pythagorean Theorem leads to a counterintuitive notion of "distance."

Finally, we can combine the lack of positive-definiteness with curvature; such geometries are called *Lorentzian*, and describe general relativity. We will study several of these geometries in subsequent chapters.

The different types of geometries referred to above are summarized in Table 1.4. But what does signature mean? In relativity, both special and general, we distinguish between *spacelike* intervals, for which $ds^2 > 0$, *timelike* intervals, for which $ds^2 < 0$, and *lightlike* (or *null*) intervals, for which $ds^2 = 0$. For spacelike intervals, ds measures the *distance* between nearby points (along a given path). For timelike intervals,

$$d\tau = \sqrt{-ds^2} \tag{1.4}$$

measures the *time* between nearby points (along a given trajectory). The parameter τ is often referred to as *proper time*, generalizing the familiar interpretation of s as *arclength*.

We begin our exploration of these geometries with a review of trigonometry.

	$s = 0$	$s = 1$
flat	Euclidean	Minkowskian
curved	Riemannian	Lorentzian

TABLE 1.4. Classification of geometries by curvature and signature.

1.2 CIRCLE TRIGONOMETRY

What is the fundamental idea in trigonometry? Although often introduced as the study of triangles, trigonometry is really the study of circles.

One way to construct the trigonometric functions is as follows:

- Draw a circle of radius r, that is, the set of points at constant distance r from the origin.
- Measure arclength s along the circle by integrating the (square root of the) line element (1.1).
- *Define* angle measure as $\phi = s/r$.
- Assuming an angle in standard position (counterclockwise from the positive x-axis), *define* the coordinates of the (other) point where the sides of the angle meet the given circle to be $(r\cos\phi, r\sin\phi)$, as shown in Figure 1.5.

Notice the key role that arclength plays in this construction. To measure an angle, and hence to define the trigonometric functions, one must know how to measure arclength.[3]

See Chapter 3 of [2] for further details.

[3]This is not as easy as it sounds. It is straightforward to use $x^2 + y^2 = r^2$ to obtain $x\,dx + y\,dy = 0$, so that $ds^2 = r^2\,dy^2/x^2$. But it is not obvious how to integrate ds without using a trigonometric substitution! This can, however, be done, since a little algebra shows that

$$i\,d\phi = \frac{i}{r}\,ds = \frac{i\,dy}{x}\,\frac{1 + i\frac{y}{x}}{1 + i\frac{y}{x}} = \frac{d(x+iy)}{x+iy}$$

which leads to *Euler's formula*,

$$r\,e^{i\phi} = x + iy,$$

relating the sine and cosine functions to the complex exponential function, and expressing arclength in terms of the complex logarithm function. Power series expansions can now be used to determine arclength as a function of position, and to provide expressions for the trigonometric functions in terms of their argument. Yes, one could instead simply define π as usual as the ratio of circumference to diameter, but that construction does not generalize to other geometries.

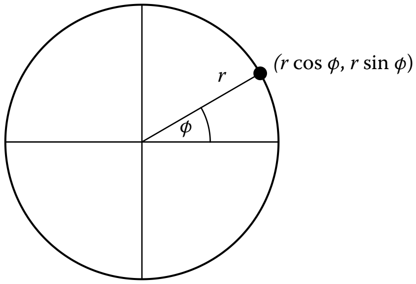


FIGURE 1.5. Defining the (circular) trigonometric functions via the unit circle.

1.3 HYPERBOLA TRIGONOMETRY

We now apply the same procedure to Lorentzian hyperbolas rather than Euclidean circles, as illustrated in Figure 1.6.

- Draw a hyperbola of "radius" ρ, that is, the set of points at constant (squared) distance $r^2 = x^2 - t^2$ from the origin.
- Measure arclength τ along the hyperbola by integrating the (square root of the absolute value of the) line element (1.3). That is, integrate $d\tau$, where $d\tau^2 = -ds^2 = dt^2 - dx^2$.
- *Define* angle measure as $\beta = \tau/\rho$.
- Assuming an angle in standard position (counterclockwise from the positive x-axis), *define* the coordinates of the (other) point where the sides of the angle meet the given hyperbola to be $(\rho\cosh\beta, \rho\sinh\beta)$.

Again, notice the key role that arclength plays in this construction.[4] See Chapter 4 of [2] for further details.

[4] And again, this is not as easy as it sounds. It is straightforward to use $x^2 - t^2 = \rho^2$ to obtain $x\,dx - t\,dt = 0$, so that $d\tau^2 = \rho^2\,dt^2/x^2$. But it is not obvious how to integrate $d\tau$ without using a (hyperbolic) trigonometric substitution! Again, some algebra helps, since

$$d\beta = \frac{1}{\rho}\,d\tau = \frac{dt}{x}\,\frac{1+\frac{t}{x}}{1+\frac{t}{x}} = \frac{d(x+t)}{x+t},$$

so that

$$\rho\,e^{\beta} = x + t$$

from which the hyperbolic trigonometric functions can be expressed as usual in terms of the exponential function, and arclength in terms of the logarithm function.

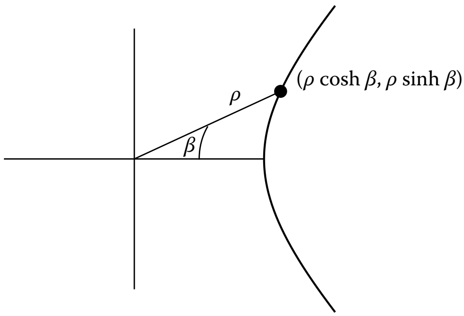


FIGURE 1.6. Defining the hyperbolic trigonometric functions via a (Lorentzian) hyperbola.

1.4 THE GEOMETRY OF SPECIAL RELATIVITY

A *spacetime diagram* in special relativity is just a diagram drawn using hyperbola geometry. Vertical lines represent the *worldline* of an observer standing still (in the given reference frame). Horizontal lines represent a "snapshot" of time, according to that observer.

Lines tilted away from the vertical axis represent objects in motion, whose speed is given by the (reciprocal) slope of the line. The basic postulate of special relativity is that the speed of light is the same for all observers. Thus, the speed of light is special. But hyperbola geometry already has a special slope, namely ± 1, corresponding to the asymptotes of all the hyperbolas. Equivalently, the two lines through the origin with slope ± 1 can be thought of as the degenerate hyperbola $x^2 - t^2 = 0$, all of whose points are at zero "distance" from the origin.

Without further ado, we henceforth adopt units such that the speed of light is 1. In other words, we measure both space and time in the same units, typically meters. Some authors indicate this by using coordinates x and ct, with t in seconds and $c = 3 \times 10^8$ meters per second; we measure time directly in meters, so that $c = 1$ (and is dimensionless!).

A useful thought experiment is to consider two train cars (reference frames), moving with respect to each other at constant speed. Suppose a lamp is turned on at the center of one of the cars when they are even with each other. (Since the speed of light is constant, it doesn't matter whether the lamp is moving.) When does the light reach the front and back of the cars? Each observer sees the light travel equal distances at the speed of

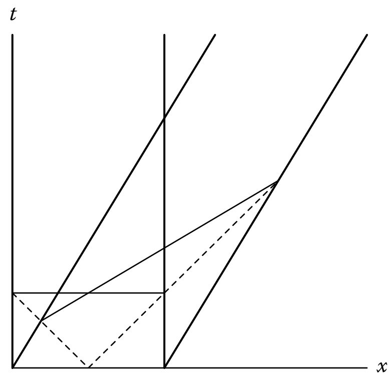


FIGURE 1.7. The heavy lines represent the ends of two train cars, one at rest (vertical lines), the other moving to the right. The dashed lines represent beams of light emitted in the center of both cars. Each observer sees the light reach both ends of their car at the same time, corresponding to two *different* notions of simultaneity ("at the same time"), as denoted by the thin lines connecting the respective intersection points.

light, so each must independently conclude that the light reaches both walls simultaneously. But each observer sees the other observer moving toward one light beam, and away from the other, and therefore concludes that the other observer does *not* see the light reach both walls simultaneously! Thus, observers in relative motion must have *different* notions of simultaneity. A spacetime diagram of this situation is shown in Figure 1.7.

See Chapter 6 of [2] for further details.

$\succ$ Chapter 2 $\prec$

Symmetries

2.1 Position and Velocity

We have given several examples of geometries and their line elements, namely the plane, the sphere, and Minkowski space. How do we describe *position* in such geometries?

The simplest method is to attach a label to the desired point, along the lines of "You are here." This method is purely geometric, but not much help in computations. One standard way to label points is to choose a coordinate system, such as rectangular coordinates, and to give the values of the chosen coordinates at the desired point. Thus, we might write

$$P = (x, y) \tag{2.1}$$

to define a particular point P in the plane with coordinates (x, y). This method works just as well when using other coordinate systems (such as polar coordinates in the plane) or on a sphere (most likely using spherical coordinates).

Another common method for labeling points is to give the *position vector* $\vec{\boldsymbol{r}}$ from the origin to the point. This works just fine in the plane, where we could write

$$\vec{\boldsymbol{r}} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}, \tag{2.2}$$

where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are unit vectors in the coordinate directions, often denoted $\hat{\boldsymbol{\imath}}$ and $\hat{\boldsymbol{\jmath}}$, respectively. But this method breaks down in general. The "position vector" on a sphere makes no sense—because there is no origin!

To fully grasp the geometric nature of relativity, it is important to understand this issue. Of course there's an origin at the center of the sphere—in $\mathbb{R}^3$. But we're not *in* $\mathbb{R}^3$, we're *on* the sphere, which is a two-dimensional surface; the radial direction does not exist. We are used to embedding such curved surfaces into (flat) $\mathbb{R}^3$ in order to visualize them, but the fact that the Earth is round can be determined using local measurements made on Earth, without leaving the planet. So trying to use the position vector to describe position is a bad idea.

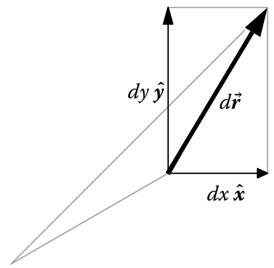


FIGURE 2.1. The rectangular components of the vector differential $d\vec{\boldsymbol{r}}$ in two dimensions, a vector form of the infinitesimal Pythagorean Theorem.

What about changes in position? Any such change is usually referred to as a *velocity*, although that label is only correct for physical objects in actual motion. Surely we can use vectors to describe changes in position?

Yes, we can; such vectors are always *in* the space being considered. So we can write

$$d\vec{\boldsymbol{r}} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}} \tag{2.3}$$

in the plane. This expression tells us that a small change in position is obtained by going a small distance (dx) in the x-direction ($\hat{\boldsymbol{x}}$), then a small distance (dy) in the y-direction ($\hat{\boldsymbol{y}}$), as shown in Figure 2.1; compare Figure 1.1. A more traditional velocity vector can be obtained from $d\vec{\boldsymbol{r}}$ by dividing by dt.

It is now easy to see how to generalize this construction to other spaces and coordinate systems. As shown in Figure 2.2, we have

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}} \tag{2.4}$$

in polar coordinates (r, ϕ), where $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\phi}}$ are the unit vectors that point in the coordinate directions. Similarly, on the sphere we have

$$d\vec{\boldsymbol{r}} = r\,d\theta\,\hat{\boldsymbol{\theta}} + r\,\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}. \tag{2.5}$$

We regard $d\vec{\boldsymbol{r}}$ as the fundamental object in vector calculus; see our online vector calculus text [1]. One of its most important properties is that

$$d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = ds^2 \tag{2.6}$$

so that $d\vec{\boldsymbol{r}}$ and ds in fact contain equivalent information about "distance." We will (almost) always assume that the basis vectors in $d\vec{\boldsymbol{r}}$ are *orthonormal*, that is, that they are mutually perpendicular unit vectors. On the

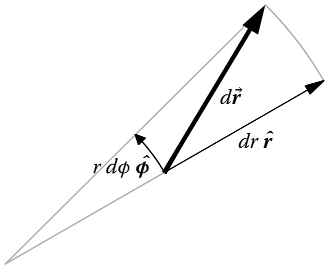


FIGURE 2.2. The components of the vector differential $d\vec{\boldsymbol{r}}$ in polar coordinates.

sphere, the basis vector $\hat{\boldsymbol{\phi}}$ points in the ϕ direction, has magnitude 1, and is orthogonal to $\hat{\boldsymbol{\theta}}$, which of course points in the θ direction (south!), and also has magnitude 1.

Minkowski space can also be described in this language. We have

$$d\vec{\boldsymbol{r}} = dx\,\hat{\boldsymbol{x}} + dt\,\hat{\boldsymbol{t}} \tag{2.7}$$

where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{t}}$ point in the x and t directions, respectively. However, $\hat{\boldsymbol{t}}$ is a *timelike* unit vector, that is

$$\hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{t}} = -1 \tag{2.8}$$

rather than $+1$.

2.2 GEODESICS

When is a curve "straight"? When its tangent vector is constant. We have[1]

$$\vec{\boldsymbol{v}} = \frac{d\vec{\boldsymbol{r}}}{d\lambda} = \dot{\vec{\boldsymbol{r}}} \tag{2.9}$$

or equivalently

$$\vec{\boldsymbol{v}}\,d\lambda = d\vec{\boldsymbol{r}} \tag{2.10}$$

from which the components of $\vec{\boldsymbol{v}}$ can be determined.

To determine whether $\vec{\boldsymbol{v}}$ is constant, we need to be able to differentiate it, and to do that we need to know how to differentiate our basis vectors, which are only constant in rectangular coordinates. For elementary

[1]We choose λ rather than t as the parameter along the curve, since t could be a coordinate. For timelike geodesics, corresponding to freely falling objects, $\vec{\boldsymbol{v}}$ is the object's *4-velocity*.

examples, such as those considered in this chapter, these derivatives can be computed by expressing each basis vector in rectangular coordinates, differentiating, and converting the answer back to the coordinate system being used. In general, however, this step requires additional information; see Chapter 17 for further details.

Once we know how to differentiate $\vec{\boldsymbol{v}}$, we can define a *geodesic* to be a curve satisfying

$$\dot{\vec{\boldsymbol{v}}} = \vec{\mathbf{0}}. \tag{2.11}$$

Since $\vec{\boldsymbol{v}}$ is itself a derivative, the geodesic equation is a set of coupled second-order ordinary differential equations in the parameter λ.

2.3 SYMMETRIES

As discussed in Section 2.2, geodesics are the solutions to a system of second-order differential equations. These equations are not always easy to solve; solving differential equations is an art form.[2] However, dramatic simplifications occur in the presence of symmetries, as we now show.

A vector $\vec{\boldsymbol{X}}$ satisfying

$$d\vec{\boldsymbol{X}} \cdot d\vec{\boldsymbol{r}} = 0 \tag{2.12}$$

is called a *Killing vector*.[3] The motivation for this definition becomes apparent if one computes

$$\frac{d}{d\lambda}(\vec{\boldsymbol{X}} \cdot \vec{\boldsymbol{v}}) = \dot{\vec{\boldsymbol{X}}} \cdot \dot{\vec{\boldsymbol{r}}} + \vec{\boldsymbol{X}} \cdot \dot{\vec{\boldsymbol{v}}}. \tag{2.13}$$

The second term in this expression vanishes if $\vec{\boldsymbol{v}}$ corresponds to a geodesic, while the first vanishes if $\vec{\boldsymbol{X}}$ is Killing. Thus, each Killing vector yields a constant of the motion along any geodesic, namely $\vec{\boldsymbol{X}} \cdot \vec{\boldsymbol{v}}$. As we will see, these constants of the motion correspond to conserved physical quantities, such as energy and angular momentum.

In what sense do Killing vectors correspond to symmetries? It turns out[4] that if the vector differential can be written in the form

$$d\vec{\boldsymbol{r}} = \sum_i h_i\, dx^i\, \hat{\boldsymbol{e}}_i, \tag{2.14}$$

[2]Some simple examples can be found in Section 19.4.

[3]Named after the German mathematician Wilhelm Killing. The dot product of vector-valued differential forms appearing in (2.12) should be thought of in the same way as in the expression $d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = ds^2$; more formally, it represents a (symmetrized) tensor product.

[4]A proof is given in Section A.1.

and if all of the coefficients h_i are independent of one of the variables, say x^j, that is, if

$$\frac{\partial h_i}{\partial x^j} = 0 \qquad\qquad (\forall i) \tag{2.15}$$

for some j, then $\vec{\boldsymbol{X}} = h_j\,\hat{\boldsymbol{e}}_j$ (no sum on j) is a Killing vector. Thus, Killing vectors correspond to directions in which $d\vec{\boldsymbol{r}}$ doesn't change.

2.4 EXAMPLE: POLAR COORDINATES

In polar coordinates, we have

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}}, \tag{2.16}$$

and the line element is

$$ds^2 = dr^2 + r^2\,d\phi^2. \tag{2.17}$$

Both of these expressions depend explicitly on r, but not ϕ. Thus, the line element does not change in the ϕ direction, and we expect "$\frac{\partial}{\partial\phi}$" to be a Killing vector. But what vector is this?

It is easy to see that

$$df = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} \tag{2.18}$$

by computing both sides of the equation in rectangular coordinates. Since coordinates do not appear explicitly in (2.18), it is a geometric statement, and must hold in any coordinate system. In fact, (2.18) can be taken as the *definition* of the gradient in curvilinear coordinates. In polar coordinates, it now follows that

$$\vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial r}\,\hat{\boldsymbol{r}} + \frac{1}{r}\,\frac{\partial f}{\partial\phi}\,\hat{\boldsymbol{\phi}} \tag{2.19}$$

so that

$$\frac{\partial f}{\partial\phi} = \vec{\boldsymbol{\nabla}} f \cdot r\,\hat{\boldsymbol{\phi}}. \tag{2.20}$$

Equivalently, to take the derivative in the ϕ direction, set $dr = 0$ and divide both sides of (2.18) by $d\phi$; then switch from ordinary derivatives to partial derivatives since we held r constant. Thus, the vector representation of $\frac{\partial}{\partial\phi}$ is the vector

$$\vec{\boldsymbol{\Phi}} = r\,\hat{\boldsymbol{\phi}}. \tag{2.21}$$

Is $\vec{\boldsymbol{\Phi}}$ really a Killing vector? After working out the derivative of the basis vector $\hat{\boldsymbol{\phi}}$, it is straightforward to compute

$$d\vec{\boldsymbol{\Phi}} = d(r\,\hat{\boldsymbol{\phi}}) = dr\,\hat{\boldsymbol{\phi}} + r\,d\hat{\boldsymbol{\phi}} = dr\,\hat{\boldsymbol{\phi}} - r\,d\phi\,\hat{\boldsymbol{r}}, \tag{2.22}$$

which is indeed orthogonal to $d\vec{\boldsymbol{r}}$ as expected, in agreement with the general result in Section A.1.

Recall now that if $\vec{\boldsymbol{X}}$ is a Killing vector and $\vec{\boldsymbol{v}}$ is (the velocity vector of) a geodesic, then $\vec{\boldsymbol{X}} \cdot \vec{\boldsymbol{v}}$ must be constant along the geodesic. Since

$$\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}} = \dot{r}\,\hat{\boldsymbol{r}} + r\,\dot{\phi}\,\hat{\boldsymbol{\phi}}, \tag{2.23}$$

from (2.9) and (2.16), we therefore have

$$\vec{\boldsymbol{\Phi}} \cdot \vec{\boldsymbol{v}} = r^2\,\dot{\phi} = \ell \tag{2.24}$$

with ℓ constant; ℓ is the constant of the motion associated with $\vec{\boldsymbol{\Phi}}$, and turns out to represent angular momentum about the origin (per unit mass). We can obtain a further condition by using arclength as the parameter along the geodesic. Dividing the line element by ds^2 yields

$$1 = \dot{r}^2 + r^2\,\dot{\phi}^2 = \dot{r}^2 + \frac{\ell^2}{r^2}. \tag{2.25}$$

We have therefore replaced the system of second-order geodesic equations (2.11) with the first-order system

$$\dot{\phi} = \frac{\ell}{r^2}, \tag{2.26}$$

$$\dot{r} = \pm\sqrt{1 - \frac{\ell^2}{r^2}}. \tag{2.27}$$

Compare this derivation with the *ad hoc* derivation given in Section 19.4, then see Section 19.5 for the explicit solution of this system of first-order equations. What do these solutions represent? Straight lines in the plane, expressed in terms of polar coordinates.

2.5 EXAMPLE: THE SPHERE

On the sphere, we have

$$d\vec{\boldsymbol{r}} = r\,d\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}, \tag{2.28}$$

and the line element is

$$ds^2 = r^2\,d\theta^2 + r^2\,\sin^2\theta\,d\phi^2. \tag{2.29}$$

Both of these expressions depend explicitly on θ, but not ϕ (or r, which is constant). Thus, the line element does not change in the ϕ direction, and we again expect "$\frac{\partial}{\partial\phi}$" to be a Killing vector.

As before, we compute

$$\vec{\boldsymbol{\nabla}} f = \frac{1}{r}\frac{\partial f}{\partial \theta}\,\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\,\frac{\partial f}{\partial \phi}\,\hat{\boldsymbol{\phi}} \tag{2.30}$$

so that

$$\frac{\partial f}{\partial \phi} = \vec{\boldsymbol{\nabla}} f \cdot r\sin\theta\,\hat{\boldsymbol{\phi}}. \tag{2.31}$$

Thus, the vector representation of $\frac{\partial}{\partial\phi}$ is the vector

$$\vec{\boldsymbol{\Phi}} = r\sin\theta\,\hat{\boldsymbol{\phi}}. \tag{2.32}$$

Is $\vec{\boldsymbol{\Phi}}$ really a Killing vector? After working out the derivative of the basis vector $\hat{\boldsymbol{\phi}}$, it is straightforward to compute

$$\begin{aligned} d\vec{\boldsymbol{\Phi}} &= d(r\sin\theta\,\hat{\boldsymbol{\phi}}) \\ &= r\cos\theta\,d\theta\,\hat{\boldsymbol{\phi}} + r\sin\theta\,d\hat{\boldsymbol{\phi}} \\ &= r\cos\theta\,d\theta\,\hat{\boldsymbol{\phi}} - r\sin\theta\cos\theta\,d\phi\,\hat{\boldsymbol{\theta}} \end{aligned} \tag{2.33}$$

which is indeed orthogonal to $d\vec{\boldsymbol{r}}$ as expected.

As before, since

$$\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}} = r\,\dot\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,\dot\phi\,\hat{\boldsymbol{\phi}}, \tag{2.34}$$

we must have

$$\vec{\boldsymbol{\Phi}} \cdot \vec{\boldsymbol{v}} = r^2\sin^2\theta\,\dot\phi = \ell \tag{2.35}$$

with ℓ constant; ℓ is the constant of the motion associated with $\vec{\boldsymbol{\Phi}}$, and turns out to represent angular momentum about the z-axis (per unit mass).

We can obtain a further condition by using arclength as the parameter along the geodesic. Dividing the line element by ds^2 yields

$$1 = r^2\dot\theta^2 + r^2\sin^2\theta\,\dot\phi^2 = r^2\dot\theta^2 + \frac{\ell^2}{r^2\sin^2\theta}. \tag{2.36}$$

We have therefore replaced the system of second-order geodesic equations (2.11) with the first-order system

$$\dot\phi = \frac{\ell}{r^2\sin^2\theta}, \tag{2.37}$$

$$\dot\theta = \pm\frac{1}{r}\sqrt{1 - \frac{\ell^2}{r^2\sin^2\theta}}. \tag{2.38}$$

Compare this derivation with the *ad hoc* derivation given in Section 19.4; then see Section 19.6 for the solution of this system of first-order equations.

What do these solutions represent? What are the "straight lines" on a sphere? Great circles, that is, diameters such as the equator, dividing the sphere into two equal halves. (Lines of latitude other than the equator are not great circles, and are *not* "straight". Among other things, they do not represent the shortest distance between two points on the sphere, as anyone who has flown across a continent or an ocean is likely aware.)

≻ Chapter 3 ≺

Schwarzschild Geometry

3.1 The Schwarzschild Metric

The Schwarzschild geometry is described by the line element

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2m}{r}} + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2. \tag{3.1}$$

As we show in Section A.10, this metric is the unique spherically symmetric solution of Einstein's equation in vacuum, and describes the gravitational field of a point mass at the origin. Yes, it also describes a black hole, but this was not realized for nearly 50 years; we will study these aspects of the Schwarzschild geometry in Chapter 5.

The geometry which bears his name was discovered by Karl Schwarzschild in early 1916 [3]—just a few months after Einstein proposed the theory of general relativity in late 1915. It was also independently discovered later in 1916 by Johannes Droste [4], a student of Hendrik Lorentz, who was the first to correctly identify the radial coordinate r in (3.1).

When writing the line element in the form (3.1), we have introduced *geometric units* not only for space (r) and time (t), but also for mass (m), *all* of which are measured in centimeters by setting both the speed of light c and the gravitational constant G to 1. If you prefer, replace t by ct, and m by mG/c^2.

The Schwarzschild metric is *asymptotically flat*, in the sense that it reduces to the Minkowski metric of special relativity as r goes to infinity. But there is clearly a problem with the Schwarzschild line element (3.1) when

$$r = 2m, \tag{3.2}$$

which is called the *Schwarzschild radius.* However, the Schwarzschild radius of the Earth is about 1 cm, and that of the Sun is about 3 km, so the peculiar effects which might occur at or near that radius are not relevant to our solar system. For example, the Earth's orbit is some 50 million Schwarzschild radii from the (center of the) Sun. We will interpret t and r as the coordinates of such a "far-away" observer, and we will assume that $r > 2m$ until further notice.

3.2 PROPERTIES OF THE SCHWARZSCHILD GEOMETRY

The basic features of the Schwarzschild geometry are readily apparent by examining the metric.

- **Asymptotic flatness:** As already discussed, the Schwarzschild geometry reduces to that of Minkowski space as r goes to infinity.

- **Spherical symmetry:** The angular dependence of the Schwarzschild metric is precisely the same as that of a sphere; the Schwarzschild geometry is *spherically symmetric*. It is therefore almost always sufficient to consider the equatorial plane $\theta = \pi/2$, so that $d\theta = 0$. For instance, a freely falling object must move in such a plane—by symmetry, it cannot deviate one way or the other. We can therefore work with the simpler line element

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2m}{r}} + r^2\, d\phi^2. \tag{3.3}$$

  As already stated, we will assume for now that $r > 2m$.

- **Circumference:** Holding r and t constant, the Schwarzschild metric reduces to that of a sphere of radius r; such surfaces are spherical shells. We can consider r to be the radius of such shells, but some care must be taken with this interpretation: The statement is *not* that there exists a ball of radius r, but only the surface of the ball. A more precise statement would be that the circumference of the spherical shell is given (at the equator) by

$$\int r\, d\phi = 2\pi r \tag{3.4}$$

  or that the area of the shell is $4\pi r^2$. Thus, r is a geometric quantity, not merely a coordinate, even though the center of the spherical shells may not exist. In this sense, "radius" refers to the dimensions of the corresponding spherical shell, rather than to "distance" from the "center."

- **Gravitational redshift:** If we consider an observer "standing still" on shells of constant radius (so $dr = 0 = d\phi$), we see that

$$d\tau = \sqrt{|ds^2|} = \sqrt{1 - \frac{2m}{r}}\, dt < dt. \tag{3.5}$$

Thus, “shell clocks” run slower than “far-away clocks”, and the period of, say, a beam of light of a given frequency will be larger for a far-away observer than for a shell observer. This is the gravitational redshift.

- **Curvature:** Similarly, if we measure the distance between two nearby shells (so $dt = 0 = d\phi$), we see that

$$ds = \frac{dr}{\sqrt{1 - \frac{2m}{r}}} > dr \tag{3.6}$$

so that the measured distance is larger than the difference in their “radii.” This is curvature.

These properties play a key role in the analysis of the Schwarzschild geometry that follows.

3.3 SCHWARZSCHILD GEODESICS

A fundamental principle of general relativity is that freely falling objects travel along timelike geodesics. So what are those geodesics?

Taking the square root of the line element, the infinitesimal vector displacement in the Schwarzschild geometry is

$$d\vec{\boldsymbol{r}} = \sqrt{1 - \frac{2m}{r}}\, dt\, \hat{\boldsymbol{t}} + \frac{dr\, \hat{\boldsymbol{r}}}{\sqrt{1 - \frac{2m}{r}}} + r\, d\theta\, \hat{\boldsymbol{\theta}} + r \sin\theta\, d\phi\, \hat{\boldsymbol{\phi}}, \tag{3.7}$$

where

$$\hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{t}} = -1, \tag{3.8}$$

$$\hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{r}} = 1 = \hat{\boldsymbol{\theta}} \cdot \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}} \tag{3.9}$$

with all other dot products between these basis vectors vanishing due to orthogonality. Spherical symmetry allows us to assume without loss of generality that any geodesic lies in the equatorial plane, so we can set $\sin\theta = 1$ and $d\theta = 0$, resulting in

$$d\vec{\boldsymbol{r}} = \sqrt{1 - \frac{2m}{r}}\, dt\, \hat{\boldsymbol{t}} + \frac{dr\, \hat{\boldsymbol{r}}}{\sqrt{1 - \frac{2m}{r}}} + r\, d\phi\, \hat{\boldsymbol{\phi}} \tag{3.10}$$

or equivalently

$$\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}} = \sqrt{1 - \frac{2m}{r}}\,\dot{t}\,\hat{\boldsymbol{t}} + \frac{\dot{r}\,\hat{\boldsymbol{r}}}{\sqrt{1 - \frac{2m}{r}}} + r\,\dot{\phi}\,\hat{\boldsymbol{\phi}}, \tag{3.11}$$

where a dot denotes differentiation with respect to proper time along the geodesic.

What other symmetries does the Schwarzschild geometry have? The line element depends on r (and θ), but not on t or ϕ. Thus, we have two Killing vectors, namely[1]

$$\vec{\boldsymbol{T}} = \sqrt{1 - \frac{2m}{r}}\,\hat{\boldsymbol{t}}, \tag{3.12}$$

$$\vec{\boldsymbol{\Phi}} = r\,\hat{\boldsymbol{\phi}}, \tag{3.13}$$

since

$$\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial t}, \tag{3.14}$$

$$\vec{\boldsymbol{\Phi}} \cdot \vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial \phi}, \tag{3.15}$$

where we have used the fact that $\theta = \pi/2$ (and that $\vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} = df$). Thus, we obtain two constants of the motion, namely

$$\ell = \vec{\boldsymbol{\Phi}} \cdot \vec{\boldsymbol{v}} = r^2\dot{\phi}, \tag{3.16}$$

$$e = -\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}} = \left(1 - \frac{2m}{r}\right)\dot{t}, \tag{3.17}$$

where $\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}}$, and where we have introduced a conventional minus sign so that $e > 0$ (for future-pointing timelike geodesics). We interpret ℓ as the angular momentum and e as the energy of the freely falling object.[2]

Our final equation is obtained from the line element itself. Since a freely falling object moves along a *timelike* geodesic, we have chosen proper time τ as our parameter. Since

$$d\tau^2 = -ds^2 = \left(1 - \frac{2m}{r}\right) dt^2 - \frac{dr^2}{1 - \frac{2m}{r}} - r^2\,d\phi^2, \tag{3.18}$$

[1]As discussed in Section 2.4 and proved in Section A.1, coordinate symmetries always lead to Killing vectors. It is straightforward to check directly that $\vec{\boldsymbol{T}}$ and $\vec{\boldsymbol{\Phi}}$ are indeed Killing vectors, but it is first necessary to determine the connection, which we postpone until Section A.3.

[2]More precisely, these expressions give the angular momentum (respectively, energy) *per unit mass* of the freely falling object. If the object has mass M, then $\vec{\boldsymbol{p}} = M\vec{\boldsymbol{v}}$ is the object's momentum, so the object's angular momentum is $L = \vec{\boldsymbol{\Phi}} \cdot \vec{\boldsymbol{p}} = M\ell$, and its energy is $E = -\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{p}} = Me$.

we can divide both sides by $d\tau^2$ to obtain

$$1 = \left(1 - \frac{2m}{r}\right) \dot{t}^2 - \frac{\dot{r}^2}{1 - \frac{2m}{r}} - r^2 \,\dot{\phi}^2, \tag{3.19}$$

where we have again used the fact that $\theta = \pi/2$.

Putting these expressions together, the (second-order) geodesic equation in the Schwarzschild geometry reduces to the (first-order) system of equations

$$\dot{\phi} = \frac{\ell}{r^2}, \tag{3.20}$$

$$\dot{t} = \frac{e}{1 - \frac{2m}{r}}, \tag{3.21}$$

$$\dot{r}^2 = e^2 - \left(1 + \frac{\ell^2}{r^2}\right)\left(1 - \frac{2m}{r}\right). \tag{3.22}$$

Before considering solutions of these equations, we first discuss the Newtonian case.

3.4 NEWTONIAN MOTION

How do objects move in Newtonian gravity? We all know that objects in orbit move on ellipses, and that "slingshot" hyperbolic trajectories also exist. The various possibilities, also including circular orbits and radial trajectories, are shown in Figure 3.1. If the source is a point mass, only radial trajectories can hit it. But where do these trajectories come from?

A standard problem in Newtonian mechanics is to analyze falling objects near the surface of the Earth, with the acceleration due to gravity assumed to be constant (g). In this case, conservation of energy takes the form

$$E = \frac{1}{2} M v^2 + M g h, \tag{3.23}$$

where the first term is the *kinetic energy* of an object of mass M traveling at speed v, and the second term is the *potential energy* of this object at height h above the surface of the Earth. But where does this expression for the potential energy come from?

The gravitational field of a point mass m is given by

$$\vec{\boldsymbol{G}} = -G\frac{m}{r^2}\,\hat{\boldsymbol{r}}, \tag{3.24}$$

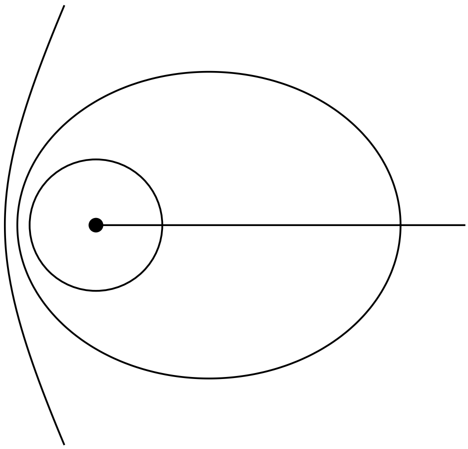


FIGURE 3.1. Newtonian particle trajectories in the gravitational field of a massive object (heavy dot) can be circular, elliptical, or hyperbolic, all with the (center of the) gravitational source at one focus. Radial trajectories are also possible; these are the only trajectories that hit the (center of the) source.

where G is the gravitational constant. The *gravitational potential* Φ satisfies

$$\vec{\boldsymbol{G}} = -\vec{\boldsymbol{\nabla}}\Phi \tag{3.25}$$

from which it follows that

$$\Phi = -G\frac{m}{r}. \tag{3.26}$$

The potential energy of an object with mass M in this gravitational field is just

$$U = M\Phi. \tag{3.27}$$

At the surface of the Earth ($r = R$), the acceleration due to gravity is

$$g = |\vec{\boldsymbol{G}}| = G\frac{m}{R^2}. \tag{3.28}$$

Finally, if we compare the potential energy at nearby radii, we get

$$U\Big|_R^{R+h} = -MG\frac{m}{R+h} + MG\frac{m}{R} = MG\frac{mh}{R(R+h)} \approx Mgh. \tag{3.29}$$

Thus, in Newtonian mechanics, the conserved energy of an object of mass M moving in the gravitational field of a point mass m is

$$E = \frac{1}{2}Mv^2 - G\frac{Mm}{r}. \tag{3.30}$$

In the equatorial plane, we have as usual

$$v^2 = \frac{ds^2}{dt^2} = \dot{r}^2 + r^2\dot{\phi}^2, \tag{3.31}$$

and we also have the conserved angular momentum[3]

$$L = Mr^2\dot{\phi}. \tag{3.32}$$

Inserting expressions (3.31) and (3.32) into (3.30), we obtain

$$\frac{1}{2}M\dot{r}^2 = E - \left(-G\frac{Mm}{r} + \frac{L^2}{2Mr^2}\right), \tag{3.33}$$

which can be rewritten in terms of the energy per unit mass ($e = E/M$) and angular momentum per unit mass ($\ell = L/M$) by dividing both sides by M.

3.5 ORBITS

In Sections 3.3 and 3.4, we have seen that the trajectory of an object moving in the gravitational field of a point mass is controlled by the object's energy per unit mass ($e = E/M$), and its angular momentum per unit mass ($\ell = L/M$). In the Newtonian case, we have

$$\frac{1}{2}\dot{r}^2 = e - \left(-\frac{m}{r} + \frac{\ell^2}{2r^2}\right) \tag{3.34}$$

from (3.33) (where we have set $G = 1$); and in the relativistic case, we have

$$\frac{1}{2}\dot{r}^2 = \frac{e^2 - 1}{2} - \left(-\frac{m}{r} + \frac{\ell^2}{2r^2} - \frac{m\ell^2}{r^3}\right) \tag{3.35}$$

from (3.22). Both of these equations take the form

$$\frac{1}{2}\dot{r}^2 = \text{constant} - V \tag{3.36}$$

where the constant depends on the energy e, and the potential V depends on both the mass m of the source and the angular momentum ℓ of the object.

[3] Why is the angular momentum proportional to r^2? For circular motion, the kinetic energy can be rewritten in terms of the angular velocity $\dot{\phi}$ as $\frac{1}{2}Mv^2 = \frac{1}{2}Mr^2\dot{\phi}^2$, which correctly suggests that the role of mass is played by the *moment of inertia* $I = Mr^2$. Thus, angular momentum is given by "mass" times "velocity", namely $L = I\dot{\phi}$.

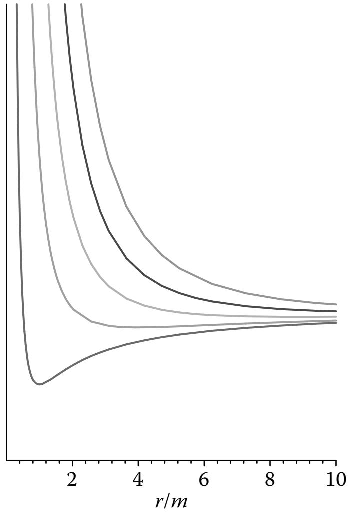


FIGURE 3.2. Newtonian potentials for various values of ℓ/m. The horizontal scale in this and subsequent potential diagrams is in multiples of the source mass m.

In the Newtonian case (3.34), the right-hand side is quadratic in r, and it is straightforward to solve this ordinary differential equation, typically by first substituting $u = 1/r$. The solutions, of course, are conic sections. In the relativistic case (3.35), the right-hand side is cubic in r. This differential equation can also be solved in closed form, but the solution requires elliptic functions as well as solving for the zeros of the cubic polynomial. Rather than present the messy details of such a computation, we give an alternative analysis here, then consider special cases such as circular and radial trajectories in subsequent sections.

A standard method for analyzing the resulting trajectories is to plot V as a function of r for various values of ℓ and m, then draw horizontal lines corresponding to the value of the constant, which depends only on the energy e. The points where these lines intersect the graph of V represent radii at which $\dot{r} = 0$; points on the line which lie above the graph have $|\dot{r}| > 0$, while points below the graph are unphysical. Plots of the potential V for various values of ℓ/m are shown in Figure 3.2 (Newtonian) and Figure 3.3 (relativistic).

A typical Newtonian example is given in Figure 3.4. The potential V has a single extremum, namely a global minimum at

$$\frac{r}{m} = \frac{\ell^2}{m^2} \tag{3.37}$$

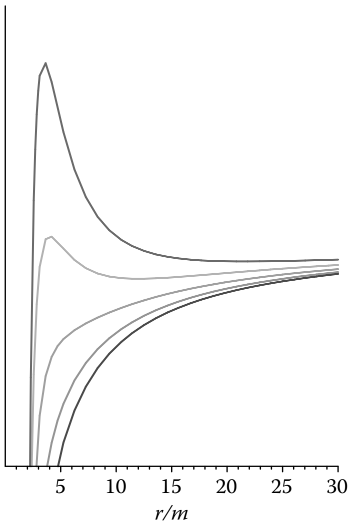


FIGURE 3.3. Relativistic potentials for various values of ℓ/m.

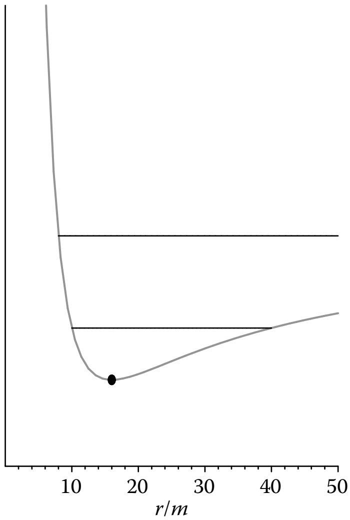


FIGURE 3.4. Newtonian potential with $\ell = 4m$.

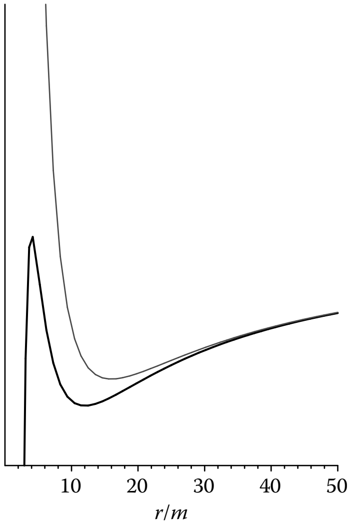


FIGURE 3.5. Newtonian and relativistic potentials with $\ell = 4m$.

which corresponds to a stable circular orbit, as indicated by the dot in Figure 3.4. With a little more energy, the object travels back and forth between a minimum radius and maximum radius; these are bound orbits, which turn out to be ellipses. These orbits are represented by the lower horizontal line in Figure 3.4. Finally, with enough energy, the object reaches a minimum radius, then turns around and never comes back; this is a "slingshot" orbit, and is represented by the upper horizontal line in Figure 3.4.[4]

Newtonian (thin line) and relativistic (heavy line) potentials with $\ell = 4m$ are shown superimposed in Figure 3.5, to emphasize that these potentials agree far from the source ($r \gg m$). However, for small r the potentials are dramatically different. In particular there is another extremum in the relativistic case, which is a local maximum, and hence corresponds to an *unstable* circular orbit. Furthermore, with enough energy it is now possible to reach $r = 0$ even with nonzero angular momentum, which is impossible in the Newtonian case.

Finally, since the condition for the location of extrema of V is quadratic in the relativistic case, there are parameter values for which no such

[4]Don't forget that the angular momentum is constant (and nonzero in the cases shown here); these objects are not falling toward the center of the source, and thus can never hit it.

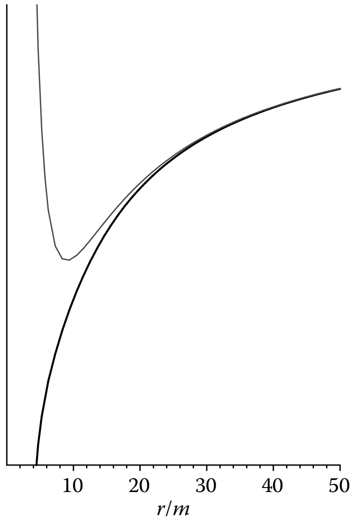


FIGURE 3.6. Newtonian and relativistic potentials with $\ell = 3m$.

extrema exist. For example, as shown in Figure 3.6, if $\ell = 3m$, the relativistic potential (heavy line) has neither minima nor maxima, and no circular orbits exist. As we will see in Section 3.6, in the relativistic case there are indeed radii at which one cannot orbit (without a boost from rocket engines), whereas in the Newtonian case one can always do so for suitable values of e and ℓ.

3.6 CIRCULAR ORBITS

Further information about the possible orbits in the Schwarzschild geometry can be obtained by analyzing the potential more carefully. Using the same conventions as before, we have

$$\frac{1}{2}\dot{r}^2 = \frac{e^2-1}{2} - V \tag{3.38}$$

with

$$V = \frac{1}{2}\left(1 - \frac{2m}{r}\right)\left(1 + \frac{\ell^2}{r^2}\right) - \frac{1}{2}. \tag{3.39}$$

Differentiating V with respect to r yields

$$V' = \frac{dV}{dr} = \frac{mr^2 - \ell^2 r + 3m\ell^2}{r^4}, \tag{3.40}$$

so that $V' = 0$ when

$$\frac{r}{m} = \frac{\ell}{2m}\left(\frac{\ell}{m} \pm \sqrt{\frac{\ell^2}{m^2} - 12}\right). \tag{3.41}$$

We have argued graphically that circular orbits can only occur when $V' = 0$, and this can only occur if

$$\ell^2 \geq 12m^2 \tag{3.42}$$

so that the roots are real. Circular orbits do not exist for all values of the parameters ℓ and e! But why must V' vanish for circular orbits?

Return to expression (3.38) for $\dot{r}$, and differentiate with respect to r, leading to

$$\dot{r}\ddot{r} = -V'\dot{r}. \tag{3.43}$$

We would like to conclude that $\ddot{r} = 0$ if and only if $V' = 0$, but this appears to require us to assume $\dot{r} \neq 0$. One could argue that well-behaved solutions must depend continuously on the parameters, so that it is legitimate to divide (3.43) by $\dot{r}$. A better argument is to begin by solving (3.38) for $\dot{r}$, yielding

$$\dot{r} = \sqrt{e^2 - 1 - 2V}, \tag{3.44}$$

so that

$$\ddot{r} = \frac{-V'\dot{r}}{\sqrt{e^2 - 1 - 2V}} = \frac{-V'\dot{r}}{\dot{r}} = -V', \tag{3.45}$$

and now the last equality can be justified in the usual way by computing the derivative using limits.

What's going on here? The use of symmetry to find constants of the motion effectively turns the second-order geodesic equation into a first-order differential equation, and this process can introduce spurious solutions. We must in principle check that our claimed solutions actually solve the full second-order equation. We will postpone that computation for now, as it first requires computing the connection in the Schwarzschild geometry, but the (r-component of the) full geodesic equation does indeed imply that

$$\ddot{r} = -V' = -\frac{mr^2 - \ell^2 r + 3m\ell^2}{r^4} \tag{3.46}$$

for *all* geodesics, including circular geodesics.[5]

[5]An analogous argument starting from Newtonian conservation of energy leads to the first-order equation

$$\frac{1}{2}M\dot{h}^2 + Mgh = \text{constant}$$

for the height h of a falling object of mass M due to the acceleration of gravity g. The fact that $h =$ constant appears to solve this equation does not imply that massive objects can hover in midair! But Newton's second law is needed to argue that $\ddot{h}$ does not vanish; the first-order equation alone is not sufficient.

Now that we have established that V' vanishes on circular orbits, it is straightforward to use (3.40) to express ℓ in terms of r, yielding

$$\frac{\ell^2}{r^2} = \frac{\frac{m}{r}}{1 - \frac{3m}{r}}, \tag{3.47}$$

and (3.38) can then be used to express e in terms of r, yielding

$$e^2 = \frac{\left(1 - \frac{2m}{r}\right)^2}{1 - \frac{3m}{r}}. \tag{3.48}$$

Not surprisingly, specifying the radius of a circular orbit completely determines the orbiting object's energy and angular momentum (per unit mass).

We can use (3.47) and (3.48) to compute the apparent angular velocity Ω of an orbiting object as seen by an observer far away, since

$$\Omega^2 = \frac{d\phi^2}{dt^2} = \frac{\dot{\phi}^2}{\dot{t}^2} = \frac{\ell^2/r^4}{e^2/\left(1 - \frac{2m}{r}\right)^2} = \frac{m}{r^3}. \tag{3.49}$$

Thus,

$$m^1 = \Omega^2 r^3, \tag{3.50}$$

which is the same "1–2–3" law as in Kepler's third law (for circular orbits).

3.7 NULL ORBITS

We repeat the computations in Section 3.6 for lightlike orbits. We still have

$$\ell = \vec{\boldsymbol{\Phi}} \cdot \vec{\boldsymbol{v}} = r^2 \dot{\phi}, \tag{3.51}$$

$$e = -\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}} = \left(1 - \frac{2m}{r}\right) \dot{t}, \tag{3.52}$$

but now we also have

$$0 = \left(1 - \frac{2m}{r}\right) \dot{t}^2 - \frac{\dot{r}^2}{1 - \frac{2m}{r}} - r^2 \, \dot{\phi}^2, \tag{3.53}$$

which leads to

$$\dot{r}^2 = e^2 - \left(1 - \frac{2m}{r}\right) \frac{\ell^2}{r^2}. \tag{3.54}$$

But wait a minute. What do these symbols mean? What does the "dot" refer to if $d\tau = 0$? And how are we supposed to interpret "energy per unit mass" (e), or "angular momentum per unit mass" (ℓ)?

To answer the first question, we note that the geodesic equation, $\dot{\vec{\boldsymbol{v}}} = 0$, automatically implies that

$$\left(|\vec{\boldsymbol{v}}|^2\right)^{\cdot} = 2\vec{\boldsymbol{v}} \cdot \dot{\vec{\boldsymbol{v}}} = 0. \tag{3.55}$$

Thus, the magnitude of the "velocity" vector $\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}}$ is constant; geodesics are curves of "constant speed." In the timelike case, we can choose the parameter to be proper time τ, and these names are reasonable. In the spacelike case, we can choose the parameter to be proper *distance*, and everything goes through as before, although the terms "velocity" and "speed" do not accurately reflect the physics. In both of these cases, it is part of the geodesic equation that we *must* choose the parameter in this way, up to a constant factor (change in units, such as between minutes and seconds) and an additive constant (shift in origin, such as between time zones). A geodesic is not merely a set of points, but rather a set of points traversed in uniform fashion. The allowable parameters are called *affine parameters*; two affine parameters u, v are related by a linear equation of the form

$$v = au + b \tag{3.56}$$

for some constants a and b.

Thus, the choice of *which* affine parameter to use is *not* formally a part of the geodesic equation, but the requirement that one use *an* affine parameter most definitely is. For timelike geodesics, we make the conventional choice to use proper time, and for spacelike geodesics we make the conventional choice to use proper distance (arclength), but for null geodesics there is no analogous preferred choice of affine parameter. The "dot" therefore refers to *any* choice of affine parameter along null geodesics.

As for the second question, this is a common problem when attempting to interpret the lightlike case as some sort of limit of timelike directions, since the "rapidity" parameter β in $\frac{v}{c} = \tanh\beta$ approaches infinity as v approaches c. The correct resolution, as shown in Figure 3.7, is that we should be looking at the *4-momentum*

$$\vec{\boldsymbol{p}} = M\,\vec{\boldsymbol{v}} \tag{3.57}$$

rather than the 4-velocity $\vec{\boldsymbol{v}}$. The quantities on the right, M and $\vec{\boldsymbol{v}}$, are separately poorly defined for lightlike objects (such as photons), but $\vec{\boldsymbol{p}}$ turns out to be a well-defined physical property of such massless objects.

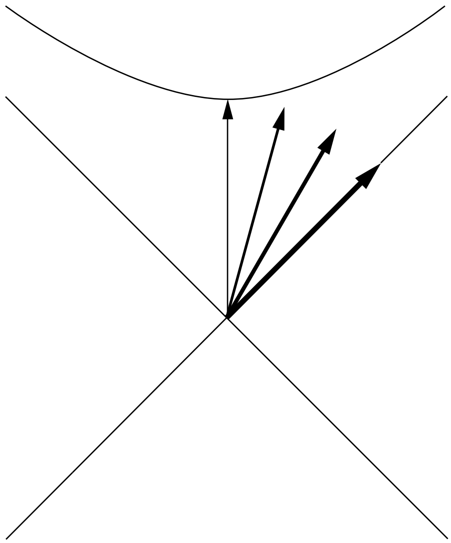


FIGURE 3.7. Lightlike momentum as a limit of timelike momenta.

Thus, for null geodesics, we would be better off defining the constants of the motion to be[6]

$$L = \vec{\mathbf{\Phi}} \cdot \vec{\boldsymbol{p}}, \tag{3.58}$$

$$E = -\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{p}}, \tag{3.59}$$

where L really is the angular momentum, and E the energy, associated with the lightlike object. We can, however, use the freedom in the affine parameter to replace e and ℓ in (3.54) by E and L, respectively.

Returning to (3.54) and differentiating, we recover the (r-component of the) full, second-order geodesic equation in the form

$$\ddot{r} = \frac{L^2}{r^3}\left(1 - \frac{3m}{r}\right), \tag{3.60}$$

and, as before, this equation still holds when $\dot{r} = 0$. Thus, the only lightlike circular orbit occurs at $r = 3m$. Equation (3.54) further implies that $L^2/E^2 = 27m^2$ on such an orbit.

[6]So why did we introduce the conserved quantities e and ℓ to begin with? Why not simply use E and L throughout? Because E and L determine different timelike geodesics for different values of M, whereas e and ℓ determine just one. In the timelike case, e and ℓ are therefore better parameters for classifying geodesics.

3.8 RADIAL GEODESICS

Consider now radial geodesics, for which $\ell = 0$, so that

$$\dot{r}^2 = e^2 - \left(1 - \frac{2m}{r}\right). \tag{3.61}$$

A radial geodesic represents a freely falling object with no angular momentum. The energy e determines the radius r_0 at which the object is at rest, since if $\dot{r} = 0$ at $r = r_0$, then

$$r_0 = \frac{2m}{1 - e^2}. \tag{3.62}$$

Since r must be nonnegative, we must have

$$e^2 \leq 1 \tag{3.63}$$

and we will assume that e is positive. If $e^2 > 1$, then $\dot{r}$ is nowhere 0.

We consider here the special case $e = 1$, which corresponds to a freely falling object that starts at rest at $r = \infty$, and for which

$$\dot{r}^2 = \frac{2m}{r} \tag{3.64}$$

so that[7]

$$\dot{r} = -\sqrt{\frac{2m}{r}}. \tag{3.65}$$

Since $e = 1$, we also have

$$\dot{t} = \frac{1}{1 - \frac{2m}{r}}. \tag{3.66}$$

Thus, a far-away observer ($r \gg m$) sees the object falling with speed

$$\frac{dr}{dt} = \frac{\dot{r}}{\dot{t}} = -\left(1 - \frac{2m}{r}\right)\sqrt{\frac{2m}{r}} \tag{3.67}$$

which approaches 0 as r approaches $2m$. What does a shell observer corresponding to $r =$ constant see?

A shell observer uses the line element to measure both distance ds (with t constant) and time $d\tau$ (with r constant). For radial motion in the

[7] If $\dot{r} > 0$, the object is moving radially outward and would eventually come to rest at $r = \infty$; we simply run the trajectory backward.

equatorial plane, both $d\theta$ and $d\phi$ are 0, so we have

$$ds = \frac{dr}{\sqrt{1-\frac{2m}{r}}}, \tag{3.68}$$

$$d\tau = \sqrt{1-\frac{2m}{r}}\,dt. \tag{3.69}$$

More generally, we have written the infinitesimal vector displacement $d\vec{\boldsymbol{r}}$ (3.7) in terms of an orthonormal basis of *vectors*, so its components represent physical distances and times. For this reason, we often give these components names, and write

$$d\vec{\boldsymbol{r}} = \sigma^t\,\hat{\boldsymbol{t}} + \sigma^r\,\hat{\boldsymbol{r}} + \sigma^\theta\,\hat{\boldsymbol{\theta}} + \sigma^\phi\,\hat{\boldsymbol{\phi}} \tag{3.70}$$

and refer to $\{\sigma^i\}$ as an *orthonormal basis of 1-forms*.[8] The shell observer therefore sees the object go past with speed

$$\frac{\sigma^r}{\sigma^t} = \frac{dr\Big/\sqrt{1-\frac{2m}{r}}}{\sqrt{1-\frac{2m}{r}}\,dt} = \frac{1}{1-\frac{2m}{r}}\frac{dr}{dt} = -\sqrt{\frac{2m}{r}}. \tag{3.71}$$

This expression approaches the speed of light (-1) as r approaches $2m$. Thus, the observer far away never sees the object cross the horizon at $r = 2m$, whereas the shell observer sees the object approach the speed of light as it approaches the horizon! What is going on here?

3.9 RAIN COORDINATES

We digress briefly to introduce "rain" coordinates adapted to freely falling observers who start from rest at $r = \infty$, each of whom moves along a geodesic as described in Section 3.8. As shown there, a shell observer sees such observers fall past them with speed

$$\tanh\beta = -\sqrt{\frac{2m}{r}}. \tag{3.72}$$

Drawing a right triangle to these proportions, as shown in Figure 3.8, or using the algebraic relationships between the hyperbolic trigonometric func-

[8]This terminology will be justified in Part III.

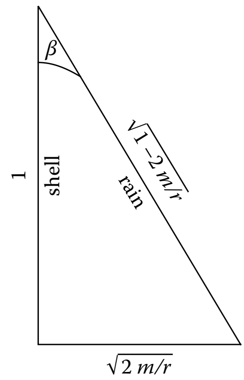


FIGURE 3.8. A hyperbolic triangle showing the hyperbolic angle between shell observers and rain observers.

tions, we find

$$\cosh\beta = \frac{1}{\sqrt{1-\frac{2m}{r}}}, \tag{3.73}$$

$$\sinh\beta = -\frac{\sqrt{\frac{2m}{r}}}{\sqrt{1-\frac{2m}{r}}}. \tag{3.74}$$

The orthonormal basis of 1-forms associated with shell observers is

$$\sigma^t = \sqrt{1-\frac{2m}{r}}\,dt, \tag{3.75}$$

$$\sigma^r = \frac{dr}{\sqrt{1-\frac{2m}{r}}}, \tag{3.76}$$

$$\sigma^\theta = r\,d\theta, \tag{3.77}$$

$$\sigma^\phi = r\,\sin\theta\,d\phi. \tag{3.78}$$

This basis is associated with shell observers, because spatial distances measured by such observers are given by integrating σ^r (with t held constant, and hence with $\sigma^t = 0$), and the time intervals measured by such observers are given by integrating σ^t (with r held constant, so that $\sigma^r = 0$).

We can obtain an orthonormal basis of 1-forms associated with freely falling observers by using a Lorentz transformation to "rotate" the shell

basis to one traveling at the correct speed. We thus define

$$\sigma^T = \sigma^t \cosh\beta - \sigma^r \sinh\beta, \tag{3.79}$$
$$\sigma^R = \sigma^r \cosh\beta - \sigma^t \sinh\beta. \tag{3.80}$$

Inserting the above expressions (3.73) and (3.74) for $\cosh\beta$ and $\sinh\beta$, we find

$$\sigma^T = dt + \frac{\sqrt{\frac{2m}{r}}}{1-\frac{2m}{r}}\,dr, \tag{3.81}$$
$$\sigma^R = \frac{dr}{1-\frac{2m}{r}} + \sqrt{\frac{2m}{r}}\,dt. \tag{3.82}$$

The first of these equations can be integrated directly, yielding

$$T = t + \int \frac{\sqrt{\frac{2m}{r}}}{1-\frac{2m}{r}}\,dr = t + 4m\left(\sqrt{\frac{r}{2m}} + \frac{1}{2}\ln\left|\frac{\sqrt{\frac{r}{2m}}-1}{\sqrt{\frac{r}{2m}}+1}\right|\right). \tag{3.83}$$

Thus,

$$\sigma^T = dT. \tag{3.84}$$

To see that this basis is indeed associated with freely falling observers, set $\sigma^R = 0$ in (3.80) ("standing still"), which implies that $\tanh\beta = \sigma^r/\sigma^t$ as desired. Furthermore, if $\sigma^R = 0$, then $\sigma^T = dT$ measures proper time. Thus, T is the proper time of the freely falling observers. We therefore refer to $\{\sigma^T, \sigma^R\}$ as the "rain" frame.[9] The relationship between rain coordinates and our original "shell" coordinates is shown in Figure 3.9.

After dividing through by $\sqrt{2m/r}$, it is also possible to integrate the second equation, resulting in an orthogonal coordinate system. However, since r has geometric meaning it is more common to retain it as a coordinate. We can nevertheless replace dt by dT in σ^R, resulting in

$$\sigma^R = dr + \sqrt{\frac{2m}{r}}\,dT. \tag{3.85}$$

[9]The "rain" coordinate system (T,r,θ,ϕ) is more properly referred to as *Painlevé–Gullstrand coordinates*; the terminology *rain coordinates* was coined in Taylor and Wheeler [5].

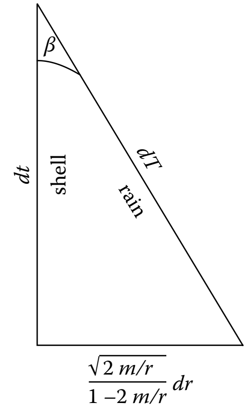


FIGURE 3.9. The relationship between shell coordinates and rain coordinates.

We can now write the line element in "rain" coordinates as

$$\begin{aligned} ds^2 &= -(\sigma^t)^2 + (\sigma^r)^2 + (\sigma^\theta)^2 + (\sigma^\phi)^2 \\ &= -(\sigma^T)^2 + (\sigma^R)^2 + (\sigma^\theta)^2 + (\sigma^\phi)^2 \\ &= -dT^2 + \left(dr + \sqrt{\frac{2m}{r}}\,dT\right)^2 + r^2\,d\theta^2 + r^2\,\sin^2\theta\,d\phi^2 \\ &= -\left(1 - \frac{2m}{r}\right)\,dT^2 + 2\sqrt{\frac{2m}{r}}\,dT\,dr \\ &\quad + dr^2 + r^2\,d\theta^2 + r^2\,\sin^2\theta\,d\phi^2. \end{aligned} \tag{3.86}$$

This line element has several unexpected properties:

- The line element is perfectly well-defined as r approaches $2m$.
- The T direction is *not* orthogonal to the r direction.
- Surfaces with T constant ($dT = 0$) are flat! (The line element reduces to that of Euclidean 3-space, in spherical coordinates.)
- The surface $r = 2m$ is lightlike! (T is a lightlike direction there.)

Since the line element is well-behaved at $r = 2m$, we can use it to ask what happens to a beam of light crossing the horizon. For a beam of light,

we have $ds^2 = 0$, and hence

$$dT^2 = \left(dr + \sqrt{\frac{2m}{r}}\, dT \right)^2, \tag{3.87}$$

so that

$$dT = \pm \left(dr + \sqrt{\frac{2m}{r}}\, dT \right); \tag{3.88}$$

or, equivalently,

$$\left(1 \mp \sqrt{\frac{2m}{r}} \right) dT = \pm dr, \tag{3.89}$$

so that

$$\frac{dr}{dT} = \pm \left(1 \mp \sqrt{\frac{2m}{r}} \right). \tag{3.90}$$

We expect $dr/dT > 0$ for an outgoing beam of light, and $dr/dT < 0$ for an ingoing beam of light. But if $r < 2m$ we have $dr/dT < 0$ in both cases! Light cannot escape from $r < 2m$! As we will see in more detail in Chapter 5, this is a black hole.

3.10 SCHWARZSCHILD OBSERVERS

An *observer* is really an army of observers stationed at all points in space. More formally, the worldlines of a family of observers foliate the spacetime. Each observer records what he or she sees, and the resulting logs are later compared. This process is often loosely described as "an observer seeing"; a better description would be "a family of observers recording."

We briefly summarize the properties of the several families of observers we have so far discussed.

- **Shell observers:** A *shell observer* stands still on a shell with r constant, so that neither θ nor ϕ is changing. Shell observers therefore measure (infinitesimal) time using

$$\sigma^t = \sqrt{1 - \frac{2m}{r}}\, dt. \tag{3.91}$$

  Shell observers measure distance using the spatial part of the Schwarzschild line element, that is, they measure distance at constant time t.

Shell observers therefore measure (infinitesimal) radial distance using

$$\sigma^r = \frac{dr}{\sqrt{1 - \frac{2m}{r}}} \tag{3.92}$$

and angular distance (in the equatorial plane, $\theta = \pi/2$) using

$$\sigma^\phi = r\,d\phi. \tag{3.93}$$

- **Freely falling observers:** Freely falling observers move along radial geodesics with $e = 1$, that is, they start at rest at $r = \infty$. They therefore measure time and radial distance using "rain" coordinates, rather than Schwarzschild coordinates, that is, using $\{\sigma^T, \sigma^R\}$ (see Section 3.9) rather than $\{\sigma^t, \sigma^r\}$. Since their motion is perpendicular to shells with $r =$ constant, they measure distance on those shells the same way shell observers do, namely with σ^ϕ.

- **Far-away "observers":** Far-away "observers" are really *bookkeepers*, rather than observers. They notice that the Schwarzschild line element reduces to the Minkowski line element for $r \gg 2m$, so that dt and dr are good approximations to "shell time" and "shell (radial) distance", respectively, so long as they are far away. But they then continue to use dt and dr to "measure" time and (radial) distance even when they are not in fact far away. One way to think of this is that far-away "observers" record the local values of the coordinates t and r, then (incorrectly) use them to compute elapsed time and distance traveled for a passing object. Thus, rather than actually measure such times and distances, they are continually "phoning home" to ask a colleague far away what time it is... However, far-away observers have no difficulty with angular motion; spherical symmetry ensures that they, too, use σ^ϕ to measure arclength for circular motion.

≻ Chapter 4 ≺

Rindler Geometry

4.1 The Rindler Metric

We digress briefly from our discussion of the Schwarzschild geometry in order to consider a much simpler geometry with many of the same properties.

Consider two-dimensional Minkowski space, with line element

$$ds^2 = -dt^2 + dx^2, \tag{4.1}$$

and introduce *hyperbolic polar coordinates* (ρ, α) via

$$x = \rho \cosh \alpha, \tag{4.2}$$

$$t = \rho \sinh \alpha, \tag{4.3}$$

so that we have

$$\rho^2 = x^2 - t^2, \tag{4.4}$$

$$\tanh \alpha = \frac{t}{x}. \tag{4.5}$$

Using these coordinates, the line element becomes

$$ds^2 = d\rho^2 - \rho^2 \, d\alpha^2, \tag{4.6}$$

which is called the *Rindler metric*; the coordinates (ρ, α) are also called *Rindler coordinates.*

Since ρ appears to be a sort of radial coordinate, and since this line element is degenerate at $\rho = 0$, we will assume $\rho > 0$; there are no restrictions on α. With these assumptions, Rindler coordinates only cover the part of Minkowski space satisfying $x > |t|$, as shown in Figure 4.1.

4.2 Properties of Rindler Geometry

What do Rindler coordinates represent? Consider a worldline with ρ constant, as shown in Figure 4.2. Since $x^2 - t^2 =$ constant, this worldline is

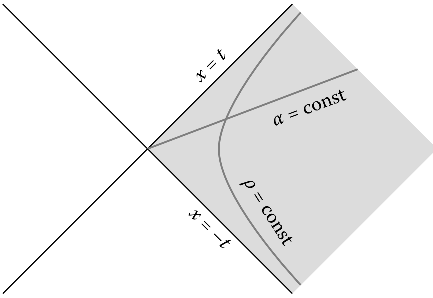


FIGURE 4.1. The shaded region is the *Rindler wedge* in Minkowski space, which is covered by the Rindler coordinates (ρ,α).

a timelike hyperbola, and is in fact one of our calibrating hyperbolas from special relativity, at constant "distance" ρ from the origin.

We claim this worldline is that of an observer undergoing constant acceleration. First of all, we have

$$v = \frac{dx}{dt} = \frac{dx/d\alpha}{dt/d\alpha} = \tanh\alpha. \tag{4.7}$$

Thus, the parameter α tells us how fast the observer is moving at any point along the trajectory, according to a standard Minkowski observer at rest. Since radial lines (α = constant) are orthogonal to such hyperbolas, such lines represent "instants of time" for our accelerating observer. To determine the acceleration, however, it is not enough to differentiate the

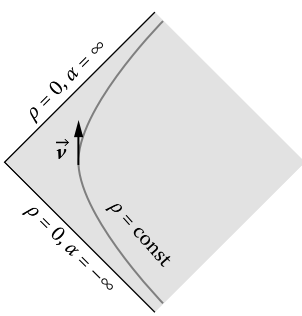


FIGURE 4.2. The 2-velocity $\vec{\boldsymbol{v}}$ of a Rindler observer, moving along a trajectory with ρ=constant.

speed v. Rather, one must differentiate the *velocity vector*, in this case the 2-velocity

$$\vec{\boldsymbol{v}} = \frac{d\vec{\boldsymbol{r}}}{d\tau} \tag{4.8}$$

as shown in Figure 4.2. In standard Minkowski coordinates (t, x), we have

$$d\vec{\boldsymbol{r}} = dt\,\hat{\boldsymbol{t}} + dx\,\hat{\boldsymbol{x}}. \tag{4.9}$$

Furthermore, proper time τ along the curve is obtained by setting $d\rho = 0$ in the line element, so that

$$d\tau = \rho\,d\alpha. \tag{4.10}$$

Thus,

$$\vec{\boldsymbol{v}} = \frac{1}{\rho}\frac{d\vec{\boldsymbol{r}}}{d\alpha} \tag{4.11}$$

which is usually written as

$$\vec{\boldsymbol{v}} = \begin{pmatrix} \dot{t} \\ \dot{x} \end{pmatrix} = \begin{pmatrix} \cosh\alpha \\ \sinh\alpha \end{pmatrix}, \tag{4.12}$$

which is clearly a unit (timelike) vector, as expected. Differentiating a second time, we obtain

$$\frac{d\vec{\boldsymbol{v}}}{d\tau} = \frac{1}{\rho}\frac{d\vec{\boldsymbol{v}}}{d\alpha} = \frac{1}{\rho}\begin{pmatrix} \sinh\alpha \\ \cosh\alpha \end{pmatrix}, \tag{4.13}$$

and the magnitude of this vector is clearly constant, namely $1/\rho$.

Thus, $\rho =$ constant curves in the Rindler geometry describe observers undergoing constant acceleration, which however depends on the "radius" ρ, and $\alpha =$ constant curves describe surfaces of constant time as seen by these observers. Within the Rindler wedge, ρ runs from 0 to ∞, while α runs from $-\infty$ to ∞.

4.3 RINDLER GEODESICS

Taking the "square root" of the line element, we have

$$d\vec{\boldsymbol{r}} = \rho\,d\alpha\,\hat{\boldsymbol{\alpha}} + d\rho\,\hat{\boldsymbol{\rho}}, \tag{4.14}$$

where

$$\hat{\boldsymbol{\alpha}} \cdot \hat{\boldsymbol{\alpha}} = -1, \tag{4.15}$$

$$\hat{\boldsymbol{\rho}} \cdot \hat{\boldsymbol{\rho}} = 1, \tag{4.16}$$

$$\hat{\boldsymbol{\alpha}} \cdot \hat{\boldsymbol{\rho}} = 0. \tag{4.17}$$

Since the line element depends only on ρ, but not on α, there is a Killing vector[1]

$$\vec{\boldsymbol{A}} = \rho\,\hat{\boldsymbol{\alpha}}, \tag{4.18}$$

since

$$\vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial \alpha} \tag{4.19}$$

(and of course $\vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} = df$). Thus, there is one constant of the motion, namely

$$e = -\vec{\boldsymbol{A}} \cdot \vec{\boldsymbol{v}} = \rho^2 \dot{\alpha}, \tag{4.20}$$

since

$$\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}} = \rho\,\dot{\alpha}\,\hat{\boldsymbol{\alpha}} + \dot{\rho}\,\hat{\boldsymbol{\rho}}. \tag{4.21}$$

Since the 2-velocity $\vec{\boldsymbol{v}}$ is always a unit vector, we also have

$$-\rho^2\,\dot{\alpha}^2 + \dot{\rho}^2 = -1, \tag{4.22}$$

which brings the geodesic equation to the form

$$\dot{\alpha} = \frac{e}{\rho^2}, \tag{4.23}$$

$$\dot{\rho}^2 = \frac{e^2}{\rho^2} - 1. \tag{4.24}$$

What does the constant e represent? Suppose that $e = 1$. Then $\dot{\rho} = 0$ when $\rho = 1$. Thus, as with radial geodesics in the Schwarzschild geometry, e determines the maximum "radius" ρ attained by an object moving along a given geodesic.

Notice that in Rindler coordinates

$$\frac{d\rho}{d\alpha} = \frac{\dot{\rho}}{\dot{\alpha}} = -\frac{\sqrt{e^2/\rho^2 - 1}}{e/\rho^2} = -\frac{\rho\sqrt{e^2 - \rho^2}}{e} \tag{4.25}$$

along a geodesic, and that this expression approaches 0 as ρ approaches 0. Thus, a single Rindler observer at fixed ρ, for whom α is (proportional to) proper time, "sees" the object stop at $\rho = 0$. On the other hand, the speed of the object is surely better described as the ratio of "Rindler distance"

[1] Again, as shown in Section A.1, coordinate symmetries always lead to Killing vectors. It is straightforward to check directly that $\vec{\boldsymbol{A}}$ is indeed a Killing vector, but it is first necessary to determine the connection, which we do not yet know how to do.

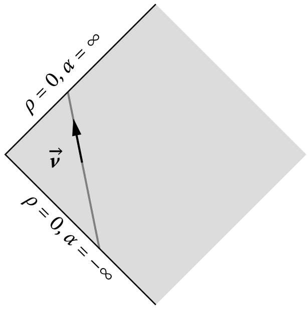


FIGURE 4.3. Rindler geodesics are straight lines in Minkowski space.

$d\rho$ to "Rindler time" $\rho\, d\alpha$, namely

$$\frac{d\rho}{\rho\, d\alpha} = -\frac{\sqrt{e^2 - \rho^2}}{e} \tag{4.26}$$

which approaches -1 as ρ approaches 0. According to this interpretation, the object is moving at the speed of light at $\rho = 0$!

Which of these interpretations is correct? Neither! We know that the geodesics in Minkowski space are straight lines, such as the one shown in Figure 4.3. An object moving along a straight line in Minkowski space will cross the line $\rho = 0$, which is just $x = \pm t$, at the same constant speed it started with.

How is this possible? A straight line approaching the future boundary $(x = t)$, as shown in Figure 4.3, rapidly approaches $\alpha = \infty$, so it is no wonder that $\frac{d\rho}{d\alpha}$ approaches 0. And (4.26) gives the *relative* speed between the object and a Rindler observer; it is the latter, not the former, that approaches the speed of light.

This example shows that the problems in the description of geodesic motion near $\rho = 0$ lie entirely with the choice of Rindler coordinates; the problems would go away if we would only switch back to Minkowski coordinates. But is there a way to see this within Rindler geometry? We address this question in Section 4.4.

4.4 EXTENDING RINDLER GEOMETRY

A useful technique in exploring an unknown geometry is to follow light beams, which leads to the use of *null coordinates*. Equivalently, factor the line element.

In Minkowski space, we have

$$ds^2 = dx^2 - dt^2 = (dx + dt)(dx - dt). \tag{4.27}$$

Thus, introduce coordinates

$$u = t - x, \tag{4.28}$$

$$v = t + x, \tag{4.29}$$

which brings the line element to the form

$$ds^2 = -du\,dv. \tag{4.30}$$

Why are null coordinates useful? Because the surfaces (curves) along which u or v is constant are lightlike.

Apply this idea to the Rindler geometry. Along a lightlike curve, we have

$$0 = ds^2 = -\rho^2\,d\alpha^2 + d\rho^2 \tag{4.31}$$

so that

$$d\alpha \pm \frac{d\rho}{\rho} = 0. \tag{4.32}$$

We therefore introduce null coordinates via

$$du = d\alpha - \frac{d\rho}{\rho}, \tag{4.33}$$

$$dv = d\alpha + \frac{d\rho}{\rho}. \tag{4.34}$$

Since we have been careful to separate variables, we can integrate these expressions to obtain

$$u = \alpha - \ln\rho, \tag{4.35}$$

$$v = \alpha + \ln\rho. \tag{4.36}$$

We must now regard ρ as a function of u and v given by

$$2\ln\rho = v - u \tag{4.37}$$

or, equivalently,

$$\rho^2 = e^{v-u}. \tag{4.38}$$

Rewriting the line element in terms of our new coordinates (u,v) therefore yields

$$ds^2 = -\rho^2\,du\,dv = -e^{v-u}\,du\,dv. \tag{4.39}$$

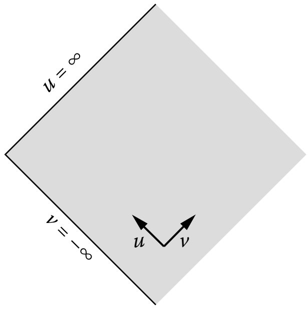


FIGURE 4.4. The null coordinates (u, v), covering (only) the Rindler wedge.

The (u, v) coordinate system is shown in Figure 4.4. By using null coordinates, we have succeeded in "straightening out" the coordinate system, but it still only covers the Rindler wedge, with both u and v running from $-\infty$ to $+\infty$.

However, the expression (4.39) for the line element clearly factors, that is,

$$ds^2 = -\left(e^{-u}\,du\right)\left(e^{v}\,dv\right) = -d\left(-e^{-u}\right)d\left(e^{v}\right). \tag{4.40}$$

We are therefore led to introduce yet another set of new coordinates

$$U = -e^{-u} = -\rho\, e^{-\alpha}, \tag{4.41}$$
$$V = e^{v} = \rho\, e^{\alpha}. \tag{4.42}$$

The rest is easy, since

$$t = \frac{V+U}{2} = \rho \sinh\alpha, \tag{4.43}$$
$$x = \frac{V-U}{2} = \rho \cosh\alpha. \tag{4.44}$$

The null coordinates (U, V) are shown in Figure 4.5, and do, indeed, cover all of Minkowski space.

We have thus recovered Minkowski coordinates, defined on all of Minkowski space, starting from Rindler coordinates, defined only on a "Rindler wedge." In the process, we have extended Rindler geometry across the "horizon" at $\rho = 0$, which turns out to be nothing more than a coordinate singularity.

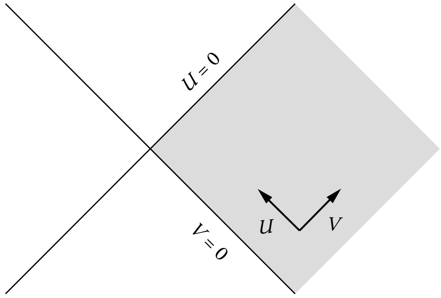


FIGURE 4.5. The null coordinates (U, V), covering all of Minkowski space.

As we will see in Chapter 5, the same process can be applied to the Schwarzschild geometry, and leads to a similar conclusion.

≻ Chapter 5 ≺

Black Holes

5.1 Extending Schwarzschild Geometry

We now apply the same technique to the Schwarzschild geometry that we used in Chapter 4 to extend the Rindler geometry. Consider a radial light beam, so that $d\phi = 0$ (and as usual $\theta = \pi/2$). Then the line element becomes

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2m}{r}} = -\left(1 - \frac{2m}{r}\right)\left(dt^2 - \frac{dr^2}{\left(1 - \frac{2m}{r}\right)^2}\right), \tag{5.1}$$

which we can factor as

$$ds^2 = -\left(1 - \frac{2m}{r}\right)\left(dt - \frac{dr}{1 - \frac{2m}{r}}\right)\left(dt + \frac{dr}{1 - \frac{2m}{r}}\right). \tag{5.2}$$

This motivates the definition

$$du = dt - \frac{dr}{1 - \frac{2m}{r}}, \tag{5.3}$$

$$dv = dt + \frac{dr}{1 - \frac{2m}{r}}, \tag{5.4}$$

which we can integrate to obtain

$$\begin{aligned} \frac{v-u}{2} &= \int \frac{dr}{1 - \frac{2m}{r}} \\ &= \int \left(1 + \frac{1}{\frac{r}{2m} - 1}\right) dr \\ &= r + 2m \ln\left(\frac{r}{2m} - 1\right). \end{aligned} \tag{5.5}$$

Expression (5.5) relating r to u and v is badly behaved at $r = 2m$. However, for $r > 2m$, the coordinates (u, v) behave much like their namesakes

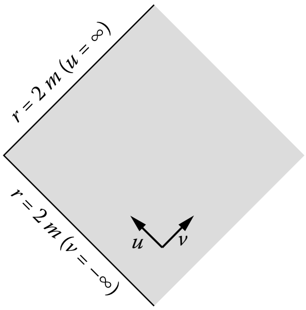


FIGURE 5.1. The null coordinates (u, v) in the Schwarzschild geometry.

in Rindler geometry, as shown in Figure 5.1 (compare Figure 4.4). Furthermore, we can extend the geometry across the apparent singularity at $r = 2m$ if we exponentiate, which yields

$$e^{(v-u)/4m} = e^{r/2m}\left(\frac{r}{2m} - 1\right) = e^{r/2m}\frac{r}{2m}\left(1 - \frac{2m}{r}\right). \tag{5.6}$$

Inserting (5.6) back into the factored line element (5.2) leads to

$$\begin{aligned} ds^2 &= -\frac{2m}{r}e^{-r/2m}e^{(v-u)/4m}\,du\,dv \\ &= -\frac{32m^3}{r}e^{-r/2m}\left(e^{-u/4m}\,\frac{du}{4m}\right)\left(e^{v/4m}\,\frac{dv}{4m}\right) \\ &= -\frac{32m^3}{r}e^{-r/2m}\,dU\,dV, \end{aligned} \tag{5.7}$$

where

$$U = -e^{-u/4m}, \tag{5.8}$$
$$V = e^{v/4m}. \tag{5.9}$$

The line element (5.7) is perfectly well behaved at $r = 2m$! The coordinates $\{U, V, \theta, \phi\}$ are known as (double-null) *Kruskal–Szekeres coordinates*. The radial coordinate r can be expressed (implicitly) in terms of U and V via

$$UV = -e^{(v-u)/4m} = e^{r/2m}\left(1 - \frac{r}{2m}\right). \tag{5.10}$$

Finally, orthogonal coordinates $\{T, X\}$ can be introduced if desired by writing

$$U = T - X, \tag{5.11}$$
$$V = T + X. \tag{5.12}$$

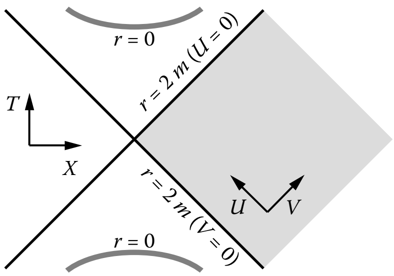


FIGURE 5.2. Kruskal geometry.

5.2 KRUSKAL GEOMETRY

Since Kruskal–Szekeres coordinates U and V are well-defined at $r = 2m$, we can use them to extend Schwarzschild geometry to $r < 2m$. The *maximally extended Schwarzschild geometry*, also called the *Kruskal geometry*, is obtained by extending the domains of U, V as much as possible, and is shown in Figure 5.2.

Since

$$UV = e^{r/2m}\left(1 - \frac{r}{2m}\right), \tag{5.13}$$

we see that $U = 0$ or $V = 0$ where $r = 2m$, and the original Schwarzschild geometry corresponds to the quadrant $U < 0$, $V > 0$. We would like to consider all possible values of U and V, but what values are allowed? Kruskal–Szekeres coordinates are badly behaved at $r = 0$, which turns out to be a true singularity, since the curvature blows up there. We therefore require $r > 0$, which implies that

$$UV < 1. \tag{5.14}$$

Thus, the U and V axes correspond to the horizons at $r = 2m$, and the singularities at $r = 0$ correspond to $UV = 1$, and are shown as thick lines. The shaded region represents the original Schwarzschild region, with $r > 2m$.

Figure 5.2 reveals the true nature of the Schwarzschild black hole. The U and V axes divide the Kruskal geometry into four regions, as shown in Figure 5.3. Since the angular coordinates θ and ϕ have been suppressed, each point in this diagram represents a sphere of radius r. The original Schwarzschild region is the righthand quadrant, which is bordered by two horizons, one in the past, and the other in the future. Since the horizons at

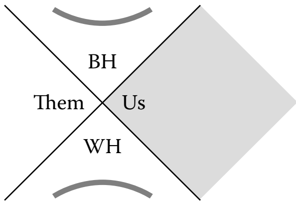


Figure 5.3. The four regions of Kruskal geometry.

$U = 0$ and $V = 0$ are lightlike, nothing can escape from the top quadrant; this region describes (the interior of a) *black hole*. Once you're in this region, you can only move upward in the diagram (forward in time), at less than the speed of light. Not even a beam of light that starts in the black hole region can escape to the outside ($r > 2m$), where we could see it! There is also a time-reversed copy of this region in the bottom quadrant, from which everything must eventually escape; this region describes a *white hole*. Finally, there is a second Schwarzschild region on the left; however, it cannot communicate with the original Schwarzschild region on the right. For this reason, the two Schwarzschild regions are often referred to as "us" and "them."

Lines of constant r (hyperbolas) and t (straight lines) are shown in Figure 5.4. This diagram should remind you of Rindler geometry! Shell observers at constant r must accelerate in order to avoid falling into the black hole; surfaces of constant t must be orthogonal to the worldlines of such observers. The same description holds inside the black hole region, with $r < 2m$, but the roles of r and t have been reversed: r is now time. Moving into the future thus forces r to decrease; it is not possible to avoid reaching the singularity at $r = 0$, and it turns out that a falling object does so in finite proper time. There is no escape from a black hole!

What does the Kruskal geometry look like at any instant of time? Since surfaces of constant t all meet at the origin, they do not represent a good global notion of "constant time." (This notion of "time" would also run backward in the left-hand quadrant!) We therefore analyze the geometry of surfaces of constant T, as shown in Figure 5.5. Start at the bottom, with the line $T = 0$. Each point on this line represents a sphere, and it is comforting to think of spheres of different radii as lying inside each other. Such a description works fine far away from the horizon, but the spheres on the left are unrelated to the spheres on the right. Imagine therefore

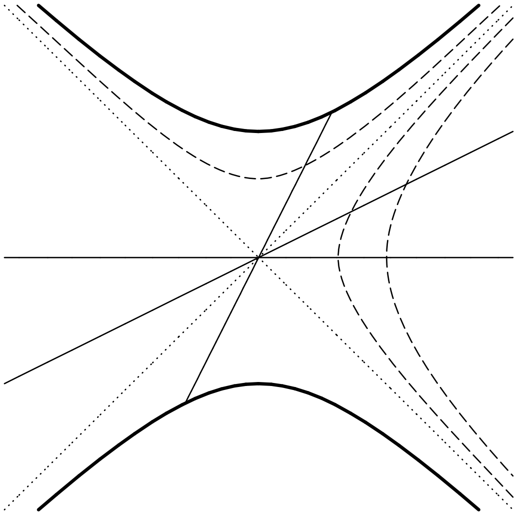


FIGURE 5.4. Lines of constant r and t in Kruskal geometry.

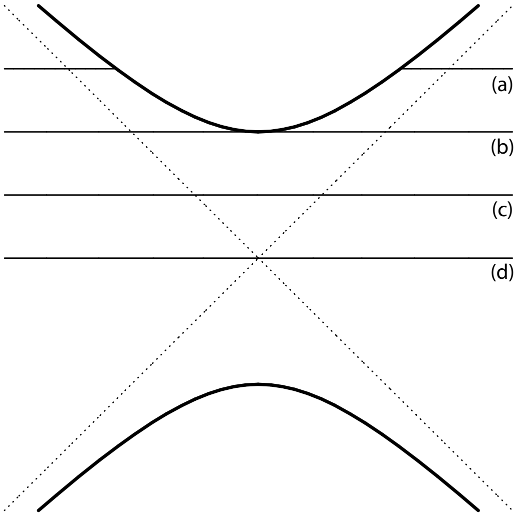


FIGURE 5.5. Lines of constant T in Kruskal geometry.

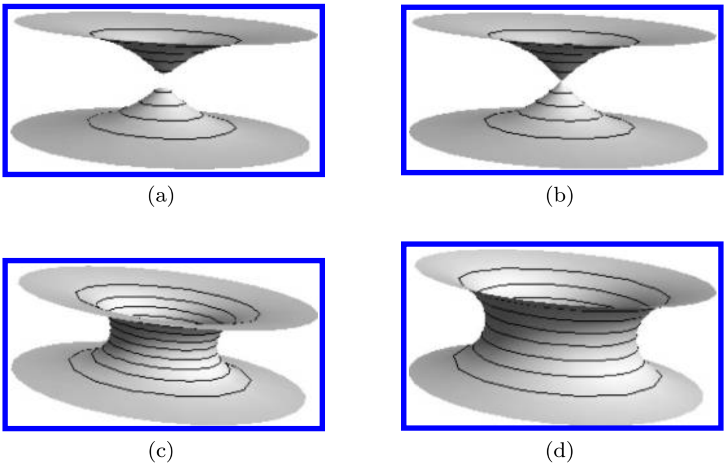


FIGURE 5.6. The topology of surfaces of constant T in the Kruskal geometry. Cases (a)–(d) correspond to the four lines shown in Figure 5.5.

two sheets of paper, on which concentric circles (representing spheres) are drawn. For $r \gg 2m$, the geometry is indeed flat. But there are no spheres on this surface with $r < 2m$, and in fact the two regions come together at $r = 2m$. This situation can be modeled by connecting the two pieces of paper along a tube, whose minimum radius is $2m$; the two Schwarzschild regions are connected by a wormhole, as shown in Figure 5.6(d). Although this figure is *drawn* in three dimensions, it must be emphasized that the vertical direction in the figure is *not* a physical direction in Kruskal geometry. Such diagrams are called *embedding diagrams* and can be quite useful for visualization, so long as one remembers that only the surface itself corresponds to (part of) Kruskal geometry.

Before incorporating such wormholes into your favorite science fiction scenario, let's explore what happens as T increases. Successive surfaces with T constant are shown in Figure 5.5, and the topology of these surfaces is shown in Figure 5.6(a)–(d). One angular coordinate is suppressed; each circle is really a sphere of constant radius. In each case, the mostly flat region at the top of the figure corresponds to the right-hand region ("us"), and the bottom to the left-hand region ("them"). As T increases (from (d) to (a)), the minimum radius on the surface decreases below $2m$. Eventually, the minimum radius of the wormhole reaches 0, after which it pinches off, after which the two Schwarzschild regions are disconnected from each other.

By following a timelike path in Kruskal geometry, it is easy to see that a Kruskal wormhole pinches off before it is possible to traverse it from one Schwarzschild region to the other; it is not possible to travel through a Kruskal wormhole.

5.3 PENROSE DIAGRAMS

A *Penrose diagram* is a spacetime diagram in which points at infinity are included. This is accomplished by rescaling the metric, that is, replacing ds^2 by $\Omega^2\, ds^2$, where the *conformal factor* Ω typically behaves like

$$\Omega \sim \frac{1}{r}. \tag{5.15}$$

Points with $\Omega = 0$ therefore correspond to points "at infinity"; more formally, this construction adds a conformal boundary to the original spacetime. Since the metric has merely been rescaled by the conformal transformation, lightlike directions are preserved, and it is customary to continue to draw such directions at 45°. Penrose diagrams are usually drawn in two dimensions, with angular degrees of freedom suppressed; a "point" in a Penrose diagram typically represents a 2-sphere. For this construction to work, the spacetime must be *asymptotically flat*, which roughly speaking means that it must look like Minkowski space "far away." All of the spacetimes we have considered so far are asymptotically flat.

The Penrose diagram for Minkowski space is shown in Figure 5.7. Angular degrees of freedom are not shown; this is the rt-plane. The dashed vertical line represents the coordinate singularity at $r = 0$; you can imagine the diagram being rotated about this axis, although each point with $r \neq 0$ corresponds to a sphere, not a circle. In Minkowski space, there are five different ways of getting infinitely far from the origin. Three are fairly obvious: far away in distance (spatial infinity, denoted i^0), and far away in time, either to the future or past (two timelike infinities, denoted $i^\pm$). Each of these infinities corresponds to a single point in the Penrose diagram, which can be thought of as 1-point compactifications, analogous to the process of mapping a plane to a sphere using stereographic projection, thus adding a single "point at infinity."

However, in Lorentzian signature, light behaves differently, and there are also two further pieces to the conformal boundary, representing the final destination of outgoing radiation and the source of incoming radiation. These boundaries are referred to as future and past null infinity,

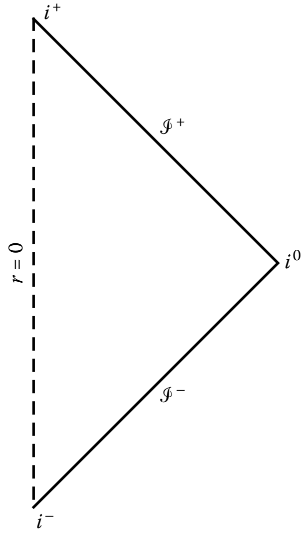


FIGURE 5.7. Penrose diagram for Minkowski space.

denoted $\mathscr{I}^\pm$, and called "Scri," for "script I." Penrose diagrams, conformal boundaries, and asymptotic flatness were all developed as tools for understanding radiation in general relativity, both electromagnetic and gravitational. Ironically, none of the early examples of asymptotically flat spacetimes, including those considered here, in fact contained any such radiation; it was only later that nontrivial asymptotically flat spacetimes were shown to exist [6–9].

The Penrose diagram for the maximally extended Schwarzschild spacetime, the Kruskal geometry, is shown in Figure 5.8. The thin lines represent the horizons at $r = 2m$, the heavy lines represent the singularities at $r = 0$, and the remaining lines represent null infinity, with $r = \infty$. As discussed in Section 5.2, the horizons divide the Kruskal geometry into four regions. The original Schwarzschild region with $r > 2m$ is the righthand asymptotic region ("us"). The upper quadrant with $r < 2m$ represents the black hole region, and the lower quadrant is a time-reversed copy, corresponding to a white hole. There is also a second asymptotic region on the left ("them"), which cannot communicate with the righthand region.

The Kruskal geometry does not correspond to any objects observed in nature; white holes do not appear to exist. A more realistic model of a

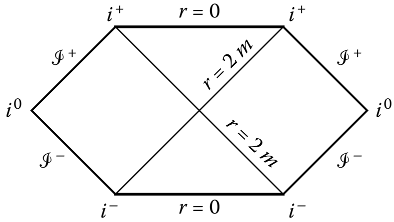


Figure 5.8. Penrose diagram for Kruskal geometry.

black hole is one created by a collapsing star, as shown in Figure 5.9. The curved line represents the surface of the star; everything to the left of this line is inside the star, and the dashed vertical line represents $r = 0$. As the star collapses, it eventually shrinks inside its Schwarzschild radius $r = 2m$, after which a black hole forms. In this model, there is no white hole region, nor is there a second asymptotic region.

Hawking has proposed that quantum mechanical processes will cause a black hole to evaporate by emitting radiation. This process is known as *Hawking radiation* and can be thought of informally as a spontaneous fluctuation in the quantum vacuum, creating a pair of particles "at" the horizon, with a positive-energy particle escaping to infinity, and a negative-energy particle falling into the black hole, thus decreasing its mass. A Penrose diagram for this model is shown in Figure 5.10. Nobody knows what happens at the point in this diagram where the black hole disappears (at the corner where $r = 0$ and $r = 2m$ intersect); the necessary theory of quantum gravity does not yet exist.

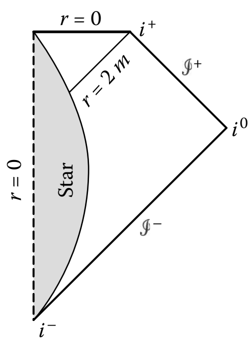


Figure 5.9. Penrose diagram for a collapsing star.

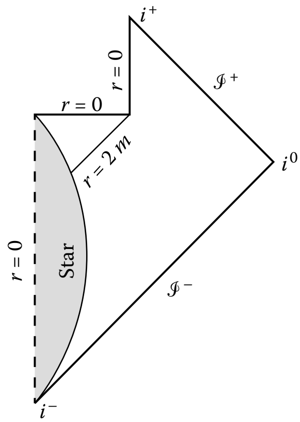


FIGURE 5.10. Penrose diagram for Hawking radiation.

5.4 CHARGED BLACK HOLES

Assuming a spherically symmetric line element of the form[1]

$$ds^2 = -f\,dt^2 + \frac{dr^2}{f} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2, \tag{5.16}$$

where f is an arbitrary function of r, Einstein's equation with an electromagnetic source, representing a point charge, can be solved for f, yielding

$$f = 1 - \frac{2m}{r} + \frac{q^2}{r^2}. \tag{5.17}$$

The resulting spacetime is known as the *Reissner–Nordström* geometry, and represents a black hole with mass m and charge q.

The global geometry of a Reissner–Nordström black hole differs considerably from that of a Schwarzschild black hole. One way to see this is to notice that (assuming $|q| < m$) there are now *two* horizons, since

$$1 - \frac{2m}{r} + \frac{q^2}{r^2} = 0 \Longrightarrow r = r_\pm = m \pm \sqrt{m^2 - q^2}, \tag{5.18}$$

where we have assumed that $m > |q| > 0$. The geometry can be extended across one horizon at a time using a procedure analogous to the Kruskal

[1]The argument used to establish Birkhoff's Theorem in Section A.10 can be generalized to support this assumption.

extension of the Schwarzschild geometry, and the resulting Penrose diagram has many asymptotic regions, not just two. The cases $m = |q| > 0$ and $|q| > m > 0$ can be handled similarly, noting that $r = m$ is a double root of f in the first case, and that there are no roots in the second case.

Further discussion, including the corresponding Penrose diagrams, can be found in Chapter 18 of d'Inverno's book [10]; see also [11].

5.5 ROTATING BLACK HOLES

The Schwarzschild geometry can be regarded as a real slice of a complex geometry, where one can perform a complex rotation and consider the resulting real slice. The resulting line element, known as the (Boyer–Lindquist form of the) *Kerr metric*, describes a rotating black hole, and is given by

$$\begin{aligned} ds^2 = -\frac{\Delta}{\rho^2}\left(dt - a\sin^2\theta\, d\phi\right)^2 + \frac{\sin^2\theta}{\rho^2}\left((r^2+a^2)\,d\phi - a\,dt\right)^2 \\ + \frac{\rho^2}{\Delta}dr^2 + \rho^2\, d\theta^2 \end{aligned} \tag{5.19}$$

where

$$\rho^2 = r^2 + a^2\cos^2\theta, \tag{5.20}$$

$$\Delta = r^2 - 2mr + a^2 \tag{5.21}$$

and where a is a constant that can be interpreted as the *angular velocity* of the black hole.

What are the properties of the Kerr geometry? First of all, it is a solution of Einstein's vacuum equation; this follows from the construction above and can also be checked directly.[2] When $a = 0$, the Kerr line element clearly reduces to the Schwarzschild line element. What symmetries does the Kerr geometry have? It is *not* spherically symmetric, but merely *axisymmetric*. There are in fact two Killing vectors, one in the ϕ direction, and the other in the t direction. There is also a discrete symmetry, namely reversing both t and ϕ, a property typical of rotation.[3]

The Kerr geometry has some new and subtle features. First of all, the singularity occurs where $\rho = 0$, which turns out to be a *ring* of radius

[2] Be careful when checking this result, as these coordinates are *not* orthogonal!

[3] The rotation leads to cross terms (containing $d\phi\, dt$) in the line element, so that the t direction is *not* orthogonal to surfaces with t = constant. There are in fact *no* hypersurfaces orthogonal to the timelike Killing vector; such Killing vectors are referred to as *stationary*, rather than *static*.

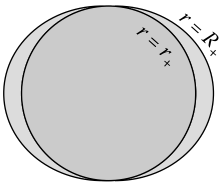


FIGURE 5.11. The ergosphere region of a Kerr black hole.

a in the equatorial plane. There are also regions with $r < 0$, and these regions contain closed timelike curves. If $|a| > m > 0$, it is even possible to pass through the ring between regions with $r < 0$ and $r > 0$, avoiding the singularity.

Perhaps even more interesting, there is now a difference between the *static limit surfaces*, where the t direction changes from timelike to spacelike, which occurs where $g_{tt} = 0$, and the horizons, where $g_{rr} \to \infty$ (more precisely, where $g^{rr} = 0$). Static limit surfaces occur where

$$\Delta - a^2 \sin^2\theta = r^2 - 2mr + a^2\cos^2\theta = 0, \tag{5.22}$$

that is, where

$$r = R_\pm = m \pm \sqrt{m^2 - a^2\cos^2\theta}, \tag{5.23}$$

whereas horizons occur where

$$\Delta = r^2 - 2mr + a^2 = 0, \tag{5.24}$$

that is, where

$$r = r_\pm = m \pm \sqrt{m^2 - a^2}. \tag{5.25}$$

As shown in Figure 5.11, there is therefore a region between the horizon at $r = r_+$ and the static limit surface at $r = R_+$, called the *ergosphere*, which is physically accessible (outside the horizon), but with the t direction spacelike. Inside the ergosphere, energy has "the wrong sign." This fact makes it possible to mine energy from a Kerr black hole by jettisoning objects from ships traveling in the ergosphere region, as originally proposed by Penrose.

Further discussion can be found in Chapter 19 of d'Inverno's book [10]; see also [11]. The above construction involving a complex rotation can also be applied to the Reissner–Nordström geometry, thus generalizing the Kerr geometry so as to include charge. The resulting geometry, known as

the *Kerr–Newman* geometry, has been shown to be the most general black hole solution possible; a black hole is completely characterized by its mass, charge, and angular velocity. In John Wheeler's famous words, "A black hole has no hair."

5.6 PROBLEMS

1. **Proper Distance between Shells**

   Recall that the Schwarzschild line element is given by

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2m}{r}} + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2.$$

   (a) Use the line element to find a formula for the proper distance between *nearby* spherical shells in the Schwarzschild geometry (surfaces with $r =$ constant). That is, find an expression for the infinitesimal distance between nearby shells, assuming that only the radius changes (and that $r > 2m$).

   (b) As you approach the horizon ($r \to 2m^+$), what happens to your expression? How far away do you think the horizon is? Do you think that you can ever get to the horizon?

2. **Approximate Proper Distance**

   The mass m of a particular black hole is 5 km, a little more than three times that of our Sun. Two concentric spherical shells surround this black hole. The inner shell has circumference $2\pi r$, and the outer shell has circumference $2\pi(r + \Delta r)$, where Δr is 100 cm. Use your expression for the infinitesimal distance between nearby shells to estimate the radial distance between the shells in each of the cases below.

   (a) $r = 50$ km

   (b) $r = 15$ km

   (c) $r = 10.5$ km

3. **Exact Proper Distance**

   (a) Use your expression for the infinitesimal distance between nearby shells to determine the exact (radial) distance traveled between two spherical shells of arbitrary circumference (but outside the horizon, that is, with $r > 2m$).

   (b) Use your result to decide whether the radial distance to the horizon is finite or infinite.

   The substitution $\cosh\alpha = \sqrt{\frac{r}{2m}}$ *may be helpful. Use your result to check the accuracy of your earlier approximations.*

4. **Satellite Orbits**

   (a) Find the speed of a satellite orbiting a Schwarzschild black hole at constant radius $r = 6m$, as measured by a stationary ("shell") observer at that radius.

   (b) Is a circular orbit at $r = \frac{5}{2}m$ possible?

   (c) Determine the smallest radius at which a circular orbit is possible, and the (shell) speed of a satellite in such an orbit.

5. **Null Orbits**
   Imagine a beam of light in orbit around a Schwarzschild black hole at constant radius.

   (a) How fast would a shell observer think the beam of light is traveling?

   (b) How fast would an observer far away think the beam of light is traveling?
   Recall that observers far away believe that t and r have their usual properties from special relativity. They are not really "observers" so much as "bookkeepers."

   (c) At what value(s) of r, if any, is such an orbit possible?

6. **Falling toward a Black Hole**

   Consider an object falling toward a Schwarzschild black hole. Assume the object starts at rest with respect to the spherical shell $r = r_0$.

   (a) According to a faraway observer, how fast does the object appear to be falling as it crosses a smaller shell whose circumference is $2\pi R$?

(b) According to an observer on that smaller shell, how fast does the object appear to be falling as it goes past?

(c) Find the limiting values of these speeds as the object approaches the horizon.
What, if anything, do you think these results mean physically? You may want to think about what happens "at" the horizon. For instance, what is the limiting value of $\frac{dt}{d\tau}$? If you "stand still" "on" the horizon, what sort of trajectory are you following? How fast does this mean you are going?

(d) How long does it take the object to reach the horizon, as measured by its own clock?

The necessary integrals are hard! Therefore assume that $r_0 = \infty$ (the object started from rest at infinity), and find the time it takes to fall from $r = R$ to $r = 2m$. You may further assume $R = 4m$ if you wish.

7. **No Escape from a Black Hole**

The goal of this problem is to establish an upper bound on the maximum proper time it takes to reach the Schwarzschild singularity at $r = 0$, starting at the horizon $r = 2m$.

(a) Use the line element to show that *inside* the horizon the following relation holds:

$$\left|\frac{dr}{d\tau}\right| \geq \sqrt{\frac{2m}{r} - 1}.$$

Assume that the Schwarzschild line element is valid inside the horizon, that is, for $r < 2m$, and recall that the proper time along any timelike trajectory is given by $d\tau^2 = -ds^2$.

(b) Assuming equality, find the elapsed proper time to travel from $r = 2m$ to $r = 0$; this is the desired upper bound.

The substitution $\sin\alpha = \sqrt{\frac{r}{2m}}$ may be helpful.

(c) How long does it take to reach the center of a solar mass black hole, starting at the horizon, if you do everything you can to resist falling in?

8. **Shining Light at a Black Hole**

   Consider lightlike, radial (both θ and ϕ constant) geodesics of the Reissner–Nordström spacetime. Show that $\ddot{r} = 0$. *This shows that r is a good affine parameter along a radial beam of light, that is, equal radial distances are covered by such a beam of light in equal "times."*

9. **Reissner–Nordström Geodesics**
   A black hole with mass m and charge q is described by the Reissner–Nordström line element

$$ds^2 = -h(r)\,dt^2 + \frac{dr^2}{h(r)} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)$$

   with $h(r) = 1 - \frac{2m}{r} + \frac{q^2}{r^2}$.

   (a) By considering the geodesic equations, or otherwise, determine a (nonzero) constant of the motion for a freely falling object moving radially in this geometry.

   (b) A massive object falls freely from rest at infinity. How fast is it moving at a given radius $r = R$ according to a shell observer at that radius?

   (c) A massive object falls freely from rest at infinity. By considering $\dot{r}$, $\ddot{r}$, or otherwise, determine the object's behavior when it approaches $r = q^2/2m$. (Assume that it does, that is, do not worry about what happens when the object reaches a horizon.)

   (d) For a solar mass black hole, whose charge is half its mass (in appropriate units), compute the value of $q^2/2m$.

   (e) Describe the object's motion. For instance, does the object ever reach $r = 0$?

10. **Null Orbits**
    Imagine a beam of light in orbit around a Reissner–Nordström black hole at constant radius. At what value(s) of r, if any, is such an orbit possible?

Part II

General Relativity

≻ Chapter 6 ≺

Warmup

6.1 Differential Forms in a Nutshell

We summarize here the basic properties of differential forms. See Chapter 12 for further details.

Wedge Products

Differential forms are integrands, the things one integrates. So dx is a differential form (a *1-form*), and so is $dx\,dy$ (a *2-form*). However, orientation matters; think about change of variables, where for instance

$$du\,dv = \frac{\partial(u,v)}{\partial(x,y)}\,dx\,dy, \tag{6.1}$$

where $\frac{\partial(u,v)}{\partial(x,y)}$ denotes the determinant of the Jacobian matrix of (u,v) with respect to (x,y). One often puts an absolute value sign around this determinant, but that's misleading. For example, flux depends on the orientation of the surface, so it's a feature, not a bug, for integrals to depend on the ordering ("handedness") of the coordinates.

In order to emphasize that order matters, we write the *exterior product* of differential forms using the symbol $\wedge$, which is read as "wedge." The Jacobian relation (6.1) becomes

$$du \wedge dv = \frac{\partial(u,v)}{\partial(x,y)}\,dx \wedge dy \tag{6.2}$$

from which we see that

$$dy \wedge dx = -dx \wedge dy, \tag{6.3}$$

$$dx \wedge dx = 0. \tag{6.4}$$

So order matters. But parentheses are not needed; the wedge product is associative (so $dx \wedge dy \wedge dz$ is unambiguous).

It is worth noting that, by construction, differential forms automatically incorporate change of variables. For example, comparing rectangular and polar coordinates, we have

$$dx \wedge dy = r\,dr \wedge d\phi, \tag{6.5}$$

an equality that does not in fact hold for differentials, despite frequent usage to the contrary.[1]

Given coordinates (x^i) on an n-dimensional space M, it is easy to see that any 1-form can be written as

$$\alpha = \alpha_i\,dx^i, \tag{6.6}$$

where we have adopted the *Einstein summation convention* in which a sum is implied by the presence of repeated indices. In other words, $\{dx^i\}$ forms a basis for 1-forms. Similarly, $\{dx^i \wedge dx^j\}$ with $i < j$ is a basis for 2-forms, and so forth. We also consider scalars (functions) to be 0-forms, and if necessary interpret wedge products of functions as ordinary multiplication.

For further insight into the meaning of $\wedge$, the reader is encouraged to work out $\alpha \wedge \beta$ for two arbitrary 1-forms in three-dimensional Euclidean space.

ORTHONORMAL FRAMES

We will (almost) always work in an orthonormal basis. Intuitively, this means that the basis 1-forms measure “infinitesimal distance” in mutually orthogonal directions. Two such bases in two-dimensional Euclidean space are $\{dx, dy\}$ and $\{dr, r\,d\phi\}$, which can easily be remembered using either the infinitesimal vector displacement

$$d\vec{\boldsymbol{r}} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}} \tag{6.7}$$

or the line element (also called the *metric*)

$$ds^2 = d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = dx^2 + dy^2 = dr^2 + r^2\,d\phi^2. \tag{6.8}$$

In general, we will write our orthonormal basis of 1-forms as $\{\sigma^i\}$, so that a generic 1-form can be expanded as

$$\beta = \beta_i\,\sigma^i. \tag{6.9}$$

[1]The correct statement in traditional language is that the corresponding *integrals* are equal. But the two notions of infinitesimal area, $dx\,dy$ and $r\,dr\,d\phi$, are *not* the same.

We introduce the notation $g(\alpha, \beta)$ for the "dot product" of two 1-forms, which is easy to compute in terms of the multiplication table for an orthonormal basis. However, we do not require the dot product to be positive definite; in particular, the "squared magnitude" of the elements in our basis can be either positive or negative. For example, in two-dimensional Minkowski space, which we have been calling hyperbola geometry and which describes special relativity, we have

$$d\vec{\boldsymbol{r}} = dt\,\hat{\boldsymbol{t}} + dx\,\hat{\boldsymbol{x}} \tag{6.10}$$

and

$$ds^2 = d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = -dt^2 + dx^2 \tag{6.11}$$

since $\hat{\boldsymbol{t}} \cdot \hat{\boldsymbol{t}} = -1$. In this geometry, we have $g(dx, dx) = 1$, but $g(dt, dt) = -1$.

It is straightforward to extend the inner product g to higher-rank differential forms, although one must be careful with signs.

HODGE DUAL

Given an orthonormal basis in n dimensions, there are exactly two choices of *volume element*, that is, of a unit n-form, obtained by multiplying together the basis 1-forms in any order, since two such products are the same if the orderings differ by an even permutation. Choosing one of these volume elements fixes an *orientation*. For example, in two Euclidean dimensions, the standard orientation is given by

$$\omega = dx \wedge dy = dr \wedge r\,d\phi. \tag{6.12}$$

Given any p-form, there is a natural $(n - p)$-form associated with it, which consists roughly of the "missing pieces" needed to make up the (given) volume element ω. We write $*\alpha$ for the differential form associated in this way with α, which is called the *Hodge dual* of α. Order matters, so $*dx = dy$, but $*dy = -dx$. In general,

$$\alpha \wedge *\beta = g(\alpha, \beta)\,\omega \tag{6.13}$$

for any p-forms α, β, and this can be used to work out $*\beta$.

An important property of the Hodge dual is that

$$** = (-1)^{p(n-p)+s} \tag{6.14}$$

where p is the rank of the differential form being acted on, and s is the *signature* of the metric, that is, the number of basis elements with negative squared magnitude.

For further insight into the meaning of $*$, the reader is encouraged to work out $*(\alpha\wedge*\beta)$ for two arbitrary 1-forms in three-dimensional Euclidean space.

EXTERIOR DIFFERENTIATION

Not only do we integrate differential forms (in the obvious way); we also differentiate them. We already know how to differentiate 0-forms, namely

$$d(f) = df = \frac{\partial f}{\partial x^i}\, dx^i. \tag{6.15}$$

We can generalize this operation to higher-rank forms by requiring

$$d(f\, dx \wedge ... \wedge dy) = df \wedge dx \wedge ... \wedge dy \tag{6.16}$$

from which it follows that

$$d^2 = 0, \tag{6.17}$$

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p\, \alpha \wedge d\beta, \tag{6.18}$$

where α is a p-form.

For further insight into the meaning of d, the reader is encouraged to work out $*d\alpha$ and $*d*\beta$ for two arbitrary 1-forms in three-dimensional Euclidean space.

CONNECTIONS AND CURVATURE

The connection 1-forms describe how the basis changes. In an orthonormal frame with

$$d\vec{\boldsymbol{r}} = \sigma^i\, \hat{\boldsymbol{e}}_i \tag{6.19}$$

the *connection 1-forms* satisfy

$$d\hat{\boldsymbol{e}}_j = \omega^i{}_j\, \hat{\boldsymbol{e}}_i, \tag{6.20}$$

$$\omega_{ij} = \hat{\boldsymbol{e}}_i \cdot d\hat{\boldsymbol{e}}_j, \tag{6.21}$$

where each $\omega^i{}_j$ and ω_{ij} differ at most by a sign, depending on the metric signature. We always work with the *Levi-Civita* connection, that is, we make the assumptions that the connection is *metric-compatible*

$$d(\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}) = d\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}} \cdot d\vec{\boldsymbol{w}} \tag{6.22}$$

and *torsion-free*

$$d^2\vec{\boldsymbol{r}} = d(d\vec{\boldsymbol{r}}) = 0. \tag{6.23}$$

For a given line element (ds^2; it is enough to give $d\vec{\boldsymbol{r}}$), there is a unique connection satisfying these conditions, which is the unique solution to the system of equations

$$d\sigma^i + \omega^i{}_j \wedge \sigma^j = 0, \tag{6.24}$$

$$\omega_{ij} + \omega_{ji} = 0, \tag{6.25}$$

where again the "up" and "down" indices reflect the signature and incorporate a factor of -1 for "negatively normed" basis elements.

The *curvature 2-forms* describe the shape of the given space and are defined by

$$d^2\hat{\boldsymbol{e}}_j = \Omega^i{}_j \hat{\boldsymbol{e}}_i, \tag{6.26}$$

which turns out to imply

$$\Omega^i{}_j = d\omega^i{}_j + \omega^i{}_m \wedge \omega^m{}_j. \tag{6.27}$$

Thus, if

$$\vec{\boldsymbol{v}} = v^i \hat{\boldsymbol{e}}_i, \tag{6.28}$$

then

$$d\vec{\boldsymbol{v}} = dv^i \hat{\boldsymbol{e}}_i + v^i d\hat{\boldsymbol{e}}_i = \left(dv^i + v^j \omega^i{}_j\right) d\hat{\boldsymbol{e}}_i \tag{6.29}$$

and

$$d^2\vec{\boldsymbol{v}} = v^i d^2\hat{\boldsymbol{e}}_i = v^i \Omega^j{}_i \hat{\boldsymbol{e}}_j. \tag{6.30}$$

Finally, we introduce the notation[2]

$$\omega^i{}_j = \Gamma^i{}_{jk}\, \sigma^k, \tag{6.31}$$

$$\Omega^i{}_j = \frac{1}{2} R^i{}_{jkm}\, \sigma^k \wedge \sigma^m \tag{6.32}$$

for the components of the connection and curvature forms; the functions $\Gamma^i{}_{jk}$ are usually referred to as *Christoffel symbols*, and the $R^i{}_{jkm}$ are the components of the *Riemann curvature tensor*.

For further insight into the meaning of connections and curvature, the reader is encouraged to compute these forms in two-dimensional Euclidean space in both rectangular and polar coordinates.

See Part III for further details.

[2] Equation (6.32) alone only determines the combinations $R^i{}_{jkm} - R^i{}_{jmk}$, so we assume by convention that $R^i{}_{jmk} = -R^i{}_{jkm}$.

6.2 TENSORS

(This section can be skipped on first reading.)

Tensors are multilinear maps on vectors. Differential forms are a type of tensor.

Consider the 1-form df for some function f. How does it act on a vector $\vec{\boldsymbol{v}}$? That's easy: by giving the directional derivative, namely

$$df(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{\nabla}} f \cdot \vec{\boldsymbol{v}}. \tag{6.33}$$

More generally, if $F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$ is a 1-form, then we define the action of F on vectors via

$$F(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{v}}. \tag{6.34}$$

There is a natural product on multilinear maps. If α and β are 1-forms, then we can construct the map

$$(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}) \longmapsto \alpha(\vec{\boldsymbol{v}})\,\beta(\vec{\boldsymbol{w}}). \tag{6.35}$$

This operation defines a new product on 1-forms, called the *tensor product*, written as[3]

$$(\alpha \otimes \beta)(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}) = \alpha(\vec{\boldsymbol{v}})\,\beta(\vec{\boldsymbol{w}}). \tag{6.36}$$

We say that $\alpha \otimes \beta$ is a *rank 2 covariant tensor*, because it is a multilinear map taking two vectors to a scalar.

Now consider the 2-form $\alpha \wedge \beta$. Since

$$g(\alpha, F) = \alpha(\vec{\boldsymbol{F}}) \tag{6.37}$$

and, as shown in Section 14.4,

$$g(\alpha \wedge \beta, \gamma \wedge \delta) = g(\alpha, \gamma) g(\beta, \delta) - g(\alpha, \delta) g(\beta, \gamma), \tag{6.38}$$

it is natural to define

$$(\alpha \wedge \beta)(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}) = \alpha(\vec{\boldsymbol{v}})\beta(\vec{\boldsymbol{w}}) - \alpha(\vec{\boldsymbol{w}})\beta(\vec{\boldsymbol{v}}). \tag{6.39}$$

With this definition, 2-forms are a special case of rank 2 covariant tensors, and in fact[4]

$$\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha. \tag{6.40}$$

[3]The parentheses around $\alpha \otimes \beta$ are usually omitted.

[4]Some authors insert factors of $n!$ into the relationship between the tensor and wedge products (with $n = 2$ in this case).

With these conventions, $(dx \wedge dy)(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) = 1$; more generally, $(dx \wedge dy)(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}})$ is the oriented area spanned by $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$.

So differential forms are *antisymmetric* tensor products of 1-forms. Another important special case is that of *symmetric* tensor products, often written as

$$\alpha \otimes_S \beta = \frac{1}{2}(\alpha \otimes \beta + \beta \otimes \alpha) \tag{6.41}$$

where the factor of $\frac{1}{2}$ ensures that

$$\alpha \otimes_S \alpha = \alpha \otimes \alpha. \tag{6.42}$$

The line element is an example of a symmetric rank 2 tensor and is often called the *metric tensor*. Terms such as dx^2 in such expressions should really be interpreted as $dx \otimes dx$.

Another example is Killing's equation, involving the dot product of vector-valued 1-forms. Such products involve a dot product on the vector-valued coefficients but should also be interpreted as implying a *symmetrized* tensor product of 1-forms. That is, we define

$$(\alpha\, \vec{\boldsymbol{u}}) \cdot (\beta\, \vec{\boldsymbol{v}}) = (\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}})\, (\alpha \otimes_S \beta), \tag{6.43}$$

resulting in a symmetric rank 2 tensor.

We will occasionally need to work with symmetric rank 2 tensors such as the metric tensor and Killing's equation. However, we only rarely need to think of such objects explicitly as multilinear maps, and will therefore omit the symbol $\otimes$. Thus, we simply write "dx^2" or "$dx\,dy$"—which is the same as "$dy\,dx$," since the order doesn't matter due to the symmetrization. It is important to remember that such products are not wedge products but can instead be manipulated using ordinary algebra.

6.3 THE PHYSICS OF GENERAL RELATIVITY

There are three key ideas underlying Einstein's theory of general relativity:

- **Principle of relativity:** The principle of relativity originates with Galileo and says that the results of experiments do not depend on relative (uniform) motion of observers. Einstein's contribution was to apply this principle to electromagnetism. Since the speed of light can be computed from the constants in Maxwell's equations, the speed of light must be independent of the observer. This postulate leads directly to special relativity.

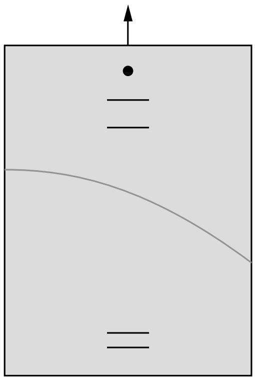


FIGURE 6.1. The behavior of light and clocks in an accelerating reference frame. The rectangle represents a rocket ship that is accelerating upward. A beam of light appears curved to observers in the ship, and clocks tick more slowly at the bottom of the ship.

- **Equivalence principle:** The equivalence principle is due to Newton and says that gravity affects everything the same way. Examples include dropping objects of different mass off the Leaning Tower of Pisa and dropping a hammer and an eagle feather on the moon. Einstein's contribution was to propose that acceleration is equivalent to gravity, so that a freely falling frame observes no (local) gravity; falling objects float in such a frame. Einstein further interpreted this to mean that freely falling observers move "straight," that is, on geodesics.[5]

- **Mach's principle:** Mach's principle is due, not surprisingly, to Mach, and says that all the matter in the universe affects the local notion of acceleration. Einstein's interpretation is that the matter in the universe curves spacetime and that this curvature then tells objects how to move under the influence of gravity. In other words, "matter = curvature," which leads to general relativity.

Simple applications of these ideas imply that gravity bends light and that clocks in a gravitational field run slow. Both of these results follow by considering an accelerating rocket ship out in space, as shown in Figure 6.1.

[5]The equivalence principle asserts a relationship between an observer moving along a curved path in a flat geometry (special relativity) and one moving along a "straight" path in a curved geometry (general relativity). The acceleration due to gravity therefore becomes the curvature of the underlying geometry, rather than the curvature of the path.

A beam of light shining through the windows that is horizontal as seen by an inertial observer follows the parabolic trajectory shown as seen by an observer on the ship. Similarly, pulses of light sent from the front (top) of the rocket to the back (bottom), indicated by the horizontal lines, will arrive more quickly than expected, since the back is accelerating to meet them, resulting in the interval between pulses being shorter upon reception than at emission. The back observer therefore thinks that the front observer's clock is running fast. Since acceleration is equivalent to gravity, both properties must also hold in a gravitational field. That is, gravity bends light, and clocks in a gravitational field tick at a slower rate than those that aren't. Equivalently, clocks in orbit tick at a faster rate than those on the ground.

The bending of light by gravity was the prediction that made Einstein famous, when this prediction was confirmed by Eddington during an eclipse in 1919. And the different rate of clocks on satellites must be taken into account by global positioning systems.

6.4 PROBLEMS

1. **Vectors in Minkowski Space**

   Show that a timelike vector cannot be orthogonal to a null vector or to another timelike vector. Show that two null vectors are orthogonal if and only if they are parallel. (Assume these vectors are nonzero.)

   Try to do this in four spacetime dimensions, rather than two. A convenient notation is to view a 4-vector $\vec{\boldsymbol{u}}$ as consisting of a time-like component u^t and spacelike components making up an ordinary 3-vector $\bar{\boldsymbol{u}}$; one often writes

   $$\vec{\boldsymbol{u}} = \begin{pmatrix} u^t \\ \bar{\boldsymbol{u}} \end{pmatrix}.$$

2. **Earth Distance**

   Corvallis, OR, is located at approximately (44.55°N,123.25°W), that is, 44.55° north of the equator (latitude), and 123.25° west of the prime meridian (longitude). Tangent, OR, is located at approximately (44.55°N,123.1°W), and Eugene, OR, is at approximately (44.05°N, 123.1°W).

Gresham, OR, is located at approximately (45.50°N,122.4°W), Millbrae, CA, is located at approximately (37.60°N,122.4°W), and Richmond, VA, is at approximately (37.60°N, 77.50°W).

Assume the Earth is a perfect sphere, with radius $r = 3959$ *miles and line element*

$$ds^2 = r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right).$$

(a) Find the approximate distance between Corvallis and Tangent.

(b) Find the approximate distance between Tangent and Eugene.

(c) Find the approximate distance between Corvallis and Eugene.

(d) Find the approximate distance between Gresham and Millbrae.

(e) Find the approximate distance between Millbrae and Richmond.

(f) Find the approximate distance between Gresham and Richmond.

(g) How good are your approximations?

3. **2-Forms**

Consider the 2-form $\tau = dx \wedge dy + dz \wedge dt$ in $\mathbb{M}^4$ (Minkowski 4-space).

(a) Find $\tau \wedge \tau$.

(b) Find $d\tau$.

(c) Either find a 1-form σ such that $d\sigma = \tau$, or argue that no such 1-form exists.

4. **Orthonormal Frames**

Consider the line element given by

$$ds^2 = -dT^2 - 2\sinh X\, dT\, dX + dX^2.$$

(a) Find an orthonormal basis of 1-forms.
The difference of squares can be factored. Can you factor ds^2*?*

(b) Determine orthogonal coordinates u, v, that is, find a coordinate transformation that brings the line element to the form $ds^2 = f\,du^2 + g\,dv^2$.

(c) What does this computation tell you about this line element?

5. **Vector Potentials**

Answer the following questions for each of the 2-forms below:

$$\beta_1 = 2yz\, dy \wedge dz + 2xz\, dz \wedge dx + 2xy\, dx \wedge dy,$$
$$\beta_2 = x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy.$$

(a) Is β_i *closed*, that is, does $d\beta_i = 0$?

(b) Is β_i *exact*, that is, does there exist a 1-form α_i such that $d\alpha_i = \beta_i$?
If β_i is not exact, explain why. If β_i is exact, find the most general solution α_i.

(c) A magnetic field $\vec{\boldsymbol{B}}$ in (Euclidean) $\mathbb{R}^3$ satisfies $\vec{\boldsymbol{B}} = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{A}}$, where $\vec{\boldsymbol{A}}$ is the vector potential. Determine which of the above differential forms, if any, could correspond to a magnetic field, and determine the corresponding vector field(s) explicitly.

$\succ$ Chapter 7 $\prec$

Geodesic Deviation

7.1 Rain Coordinates II

Recall from Section 3.9 that

$$\sigma^T = dt + \frac{\sqrt{\frac{2m}{r}}}{1 - \frac{2m}{r}}\, dr, \tag{7.1}$$

$$\sigma^R = \frac{dr}{1 - \frac{2m}{r}} + \sqrt{\frac{2m}{r}}\, dt, \tag{7.2}$$

which allows us to introduce rain coordinates (T,R) defined by

$$dT = dt + \sqrt{\frac{2m}{r}} \frac{dr}{1 - 2m/r}, \tag{7.3}$$

$$dR = \sqrt{\frac{r}{2m}} \frac{dr}{1 - 2m/r} + dt. \tag{7.4}$$

The Schwarzschild line element in rain coordinates then takes the form

$$ds^2 = -dT^2 + \frac{2m}{r} dR^2 + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right), \tag{7.5}$$

and a straightforward but lengthy computation (see Section A.4) shows that the (independent, nonzero) curvature 2-forms in these coordinates are

$$\Omega^T{}_R = \frac{2m}{r^3}\, \sigma^T \wedge \sigma^R, \tag{7.6}$$

$$\Omega^T{}_\theta = -\frac{m}{r^3}\, \sigma^T \wedge \sigma^\theta, \tag{7.7}$$

$$\Omega^T{}_\phi = -\frac{m}{r^3}\, \sigma^T \wedge \sigma^\phi, \tag{7.8}$$

$$\Omega^R{}_\theta = -\frac{m}{r^3}\, \sigma^R \wedge \sigma^\theta, \tag{7.9}$$

$$\Omega^R{}_\phi = -\frac{m}{r^3}\, \sigma^R \wedge \sigma^\phi, \tag{7.10}$$

$$\Omega^\theta{}_\phi = \frac{2m}{r^3}\, \sigma^\theta \wedge \sigma^\phi. \tag{7.11}$$

But what is the physical meaning of these expressions?

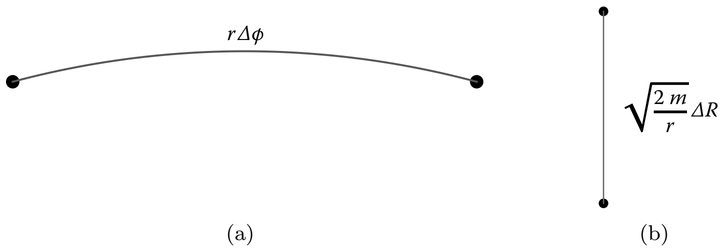


FIGURE 7.1. Nearby objects falling radially, either (a) next to each other (at equal radius) or (b) directly on top of each other.

7.2 TIDAL FORCES

Consider two nearby objects falling radially. We first consider the case where both objects start at rest on the same $r =$ constant shell, so that they have the same energy e. This situation is shown in Figure 7.1(a). The separation between the objects is therefore

$$\Delta s = r\,\Delta\phi \tag{7.12}$$

and spherical symmetry implies that

$$\Delta\phi = \text{constant.} \tag{7.13}$$

From Section 3.8 we know that both trajectories satisfy

$$\dot{r}^2 = e^2 - \left(1 - \frac{2m}{r}\right), \tag{7.14}$$

which we can differentiate to obtain

$$\ddot{r} = -\frac{m}{r^2} \tag{7.15}$$

since $\dot{r} \neq 0$. Thus,

$$(\Delta s)^{\cdot\cdot} = (r\,\Delta\phi)^{\cdot\cdot} = \ddot{r}\,\Delta\phi = -\frac{m}{r^2}\,\Delta\phi = -\frac{m}{r^3}\,\Delta s. \tag{7.16}$$

As we will show in Section 7.3, the coefficient of Δs is a component of curvature. Which one? We are taking two derivatives along the geodesic (that is, in the T-direction) of a displacement in the ϕ-direction, which is changing only in the ϕ-direction. We claim that

$$(r\,\Delta\phi\,\hat{\boldsymbol{\phi}})^{\cdot\cdot} = -R^{\phi}{}_{T\phi T}\,\hat{\boldsymbol{\phi}}, \tag{7.17}$$

which is true since

$$\Omega^{\phi}{}_{T} = \Omega^{T}{}_{\phi} = -\frac{m}{r^3}\,\sigma^T \wedge \sigma^\phi \tag{7.18}$$

(and since $\hat{\boldsymbol{\phi}}$ does not change in the r-direction).

What about two objects falling along the same trajectory, that is, with the same value of ϕ, but different values of r? This situation is shown in Figure 7.1(b). According to the metric in rain coordinates, the separation between the objects is now

$$\Delta s = \sqrt{\frac{2m}{r}}\,\Delta R. \tag{7.19}$$

Assume further that both objects start from rest at infinity, but at different times. Both trajectories therefore correspond to rain observers, for whom the rain coordinate R is constant! Thus, R is constant for each object, and therefore ΔR remains constant. (This is why rain coordinates are used for this analysis.)

We now have $e = 1$, so that

$$\dot{r} = -\sqrt{\frac{2m}{r}} \tag{7.20}$$

and

$$(\Delta s)^{\cdot\cdot} = \left(\sqrt{\frac{2m}{r}}\,\Delta R\right)^{\cdot\cdot} = \frac{2m}{r^3}\Delta s. \tag{7.21}$$

Since our separation is now measured in the R-direction, we should expect

$$\left(\sqrt{\frac{2m}{r}}\,\Delta R\,\hat{\boldsymbol{R}}\right)^{\cdot\cdot} = -R^{R}{}_{TRT}\,\hat{\boldsymbol{R}}, \tag{7.22}$$

which is true since

$$\Omega^{R}{}_{T} = \Omega^{T}{}_{R} = \frac{2m}{r^3}\,\sigma^T \wedge \sigma^R \tag{7.23}$$

(and since $\hat{\boldsymbol{R}}$ does not change in the R-direction).

Notice that falling objects next to each other accelerate *toward* each other, whereas falling objects above each other accelerate *away from* each other, just as in the Newtonian case. This shows directly how curvature influences geodesic motion, and supports the interpretation that gravity is curvature.

Equation (7.15) is in fact identical to the Newtonian expression for the acceleration due to gravity for a point mass; don't forget that we have set $G = 1$. Expressions (7.16) and (7.21) are therefore formally the same as in

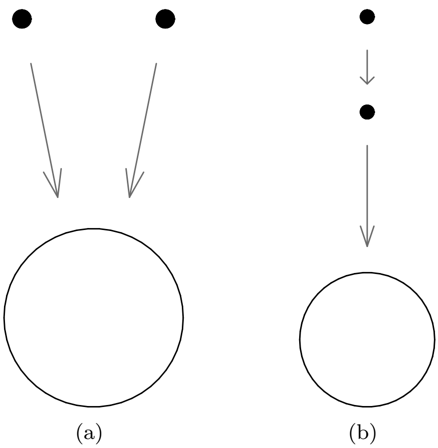


FIGURE 7.2. Two objects falling toward the Earth from far away move either (a) closer together or (b) further apart depending on their initial configuration.

Newtonian theory! To see this, note first of all that the argument leading to (7.16) is unchanged, as it relies only on the expression for $\ddot{r}$. As for objects falling on top of each other, the relative acceleration is just the difference of the respective accelerations due to gravity, so that

$$(\Delta r)^{\cdot\cdot} \approx \frac{d\ddot{r}}{dr}\Delta r = \frac{2m}{r^3}\Delta r. \tag{7.24}$$

Both of these effects are illustrated in Figure 7.2 for objects falling toward the Earth.

These effects cause tides! To see this, imagine that the Earth is falling toward the moon, as shown in Figure 7.3. The Earth itself can be regarded as a rigid body, falling toward the moon with the velocity given by the arrow at its center. But the water on the surface of the Earth falls toward the moon depending on its location, with velocities given by the four

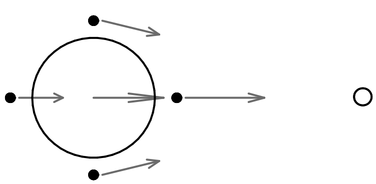


FIGURE 7.3. Tides are caused by the Earth falling toward the moon!

arrows shown. The relative motion of the water and the Earth is given by comparing the arrows: *Both* the nearest and furthest points from the moon move away from the center, causing high tides, whereas the points on either side move toward the center, causing low tides. This explains why there are two, not one, high (and low) tides every day! A similar effect, but roughly half as strong, is caused by the sun.

In both Einstein's theory and Newton's, these results demonstrate the quadrupole nature of gravity; it stretches things in one direction but shrinks them in the other.

7.3 GEODESIC DEVIATION

Consider a family of geodesics, with velocity vectors of the form

$$\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}} = \frac{\partial \vec{\boldsymbol{r}}}{\partial \tau} \tag{7.25}$$

as shown in Figure 7.4. Since these curves are geodesics, their velocity vectors satisfy

$$\dot{\vec{\boldsymbol{v}}} = 0. \tag{7.26}$$

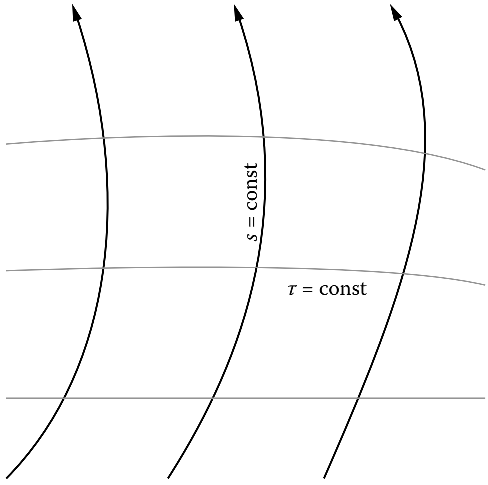


FIGURE 7.4. A family of geodesics.

Label the geodesics by some coordinate(s) s, and consider the *separation vector*

$$\vec{\boldsymbol{u}} = \vec{\boldsymbol{r}}' = \frac{\partial \vec{\boldsymbol{r}}}{\partial s} \tag{7.27}$$

between nearby geodesics. By construction, we have

$$\dot{\vec{\boldsymbol{u}}} = \vec{\boldsymbol{v}}'. \tag{7.28}$$

We are interested in the behavior of the separation $\vec{\boldsymbol{u}}$ along the geodesics, and we have

$$\ddot{\vec{\boldsymbol{u}}} = \left(\vec{\boldsymbol{v}}'\right)^{\cdot} = \left(\vec{\boldsymbol{v}}'\right)^{\cdot} - (\dot{\vec{\boldsymbol{v}}})', \tag{7.29}$$

where we have used the fact that $\vec{\boldsymbol{v}}$ is a geodesic in the last equality. This expression involves second derivatives of $\vec{\boldsymbol{v}}$ in two different directions and changes sign if those directions are interchanged. It should not be surprising that this combination of derivatives is related to $d^2\vec{\boldsymbol{v}}$. As shown in Section A.2, it turns out that

$$\ddot{\vec{\boldsymbol{u}}} = -d^2\vec{\boldsymbol{v}}(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}) = -\Omega^i{}_j(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}})v^j\,\hat{\boldsymbol{e}}_i = -R^i{}_{jk\ell}v^j u^k v^\ell\,\hat{\boldsymbol{e}}_i \tag{7.30}$$

and we can think of this as

$$\ddot{\vec{\boldsymbol{u}}} = -R^i{}_{\tau s \tau}\hat{\boldsymbol{e}}_i. \tag{7.31}$$

The s-component of this expression is indeed $-R^s{}_{\tau s \tau}$, as claimed in the examples in Section 7.2.

7.4 SCHWARZSCHILD CONNECTION

We have so far related curvature to geodesic deviation by working in rain coordinates, a special coordinate system adapted to the radial geodesics we chose to study. But this relationship is geometric, and therefore independent of the coordinates we choose. To demonstrate this geometric invariance, we revisit the problem of geodesic deviation using our original Schwarzschild coordinates. First, we need the connection.

A straightforward but lengthy computation (see Section A.3) shows that the (independent, nonzero) curvature 2-forms in these coordinates take the

form

$$\Omega^t{}_r = \frac{2m}{r^3}\,\sigma^t \wedge \sigma^r, \tag{7.32}$$

$$\Omega^t{}_\theta = -\frac{m}{r^3}\,\sigma^t \wedge \sigma^\theta, \tag{7.33}$$

$$\Omega^t{}_\phi = -\frac{m}{r^3}\,\sigma^t \wedge \sigma^\phi, \tag{7.34}$$

$$\Omega^r{}_\theta = -\frac{m}{r^3}\,\sigma^r \wedge \sigma^\theta, \tag{7.35}$$

$$\Omega^r{}_\phi = -\frac{m}{r^3}\,\sigma^r \wedge \sigma^\phi, \tag{7.36}$$

$$\Omega^\theta{}_\phi = \frac{2m}{r^3}\,\sigma^\theta \wedge \sigma^\phi. \tag{7.37}$$

Remarkably, these expressions are formally the same as the corresponding expressions in rain coordinates.

7.5 TIDAL FORCES REVISITED

The components $R^i{}_{jk\ell}$ of the curvature 2-forms form a *tensor*, known as the *Riemann curvature tensor*. Tensor components can be computed in any basis, then converted to any other basis using the appropriate change-of-basis transformations. Put differently, tensors are linear maps on vectors and can therefore easily be evaluated on any vectors, not just on basis vectors.

As an example, we recompute the tidal forces on neighboring falling objects in the Schwarzschild geometry, but this time work in Schwarzschild coordinates rather than rain coordinates. The discussion in Section 7.3 leads to

$$\ddot{\vec{\boldsymbol{u}}} = -R^s{}_{\tau s\tau}\,\vec{\boldsymbol{u}} \tag{7.38}$$

and it remains to compute $-R^s{}_{\tau s\tau}$. Assuming that the s direction is spatial, so that the up and down indices don't matter, we can think of this expression as the invariant object "$-R(\hat{s},\hat{\tau},\hat{s},\hat{\tau})$," where $\hat{s}$ and $\hat{\tau}$ denote the unit vectors in the s and τ directions, and where R is linear in each argument. But the τ direction corresponds to

$$\hat{\boldsymbol{T}} = \frac{1}{\sqrt{1-2m/r}}\left(\hat{\boldsymbol{t}} - \sqrt{\frac{2m}{r}}\,\hat{\boldsymbol{r}}\right) \tag{7.39}$$

so that by linearity we must have

$$R^s{}_{\tau s\tau} = \frac{1}{1-2m/r}R^s{}_{tst} + \frac{2m/r}{1-2m/r}R^s{}_{rsr}. \tag{7.40}$$

If the s direction corresponds to $\hat{\boldsymbol{\phi}}$, we obtain

$$\begin{aligned} R^{\phi}{}_{\tau\phi\tau} &= \frac{1}{1-2m/r} R^{\phi}{}_{t\phi t} + \frac{2m/r}{1-2m/r} R^{\phi}{}_{r\phi r} \\ &= \frac{1}{1-2m/r}\frac{m}{r^3} - \frac{2m/r}{1-2m/r}\frac{m}{r^3} = \frac{m}{r^3} \end{aligned} \tag{7.41}$$

and if the s direction corresponds to

$$\hat{\boldsymbol{R}} = \frac{1}{\sqrt{1-2m/r}}\left(\hat{\boldsymbol{r}} - \sqrt{\frac{2m}{r}}\,\hat{\boldsymbol{t}}\right) \tag{7.42}$$

then

$$\begin{aligned} R^{R}{}_{\tau R\tau} &= \frac{1}{1-2m/r} R^{r}{}_{trt} + \frac{2m/r}{1-2m/r} R^{t}{}_{rtr} \\ &= -\frac{1}{1-2m/r}\frac{2m}{r^3} + \frac{2m/r}{1-2m/r}\frac{2m}{r^3} = -\frac{2m}{r^3} \end{aligned} \tag{7.43}$$

(since terms involving $R^{t}{}_{ttt}$ or $R^{r}{}_{rrr}$ vanish).

These expressions agree with those previously computed in rain coordinates in Section 7.2.

≻ CHAPTER 8 ≺

EINSTEIN'S EQUATION

8.1 MATTER

The 2-momentum of an object in (two-dimensional) Minkowski space moving at speed $v = \tanh\beta$ is given by

$$\boldsymbol{p} = m\,\boldsymbol{u} = \begin{pmatrix} E \\ p \end{pmatrix} = \begin{pmatrix} m\,\cosh\beta \\ m\,\sinh\beta \end{pmatrix}. \tag{8.1}$$

Consider N equally spaced identical particles, as shown in Figure 8.1, and in a spacetime diagram in Figure 8.2. If the separation between particles in their rest frame is ℓ, then the *number density* of particles in their rest frame is

$$n = \frac{N}{\ell}. \tag{8.2}$$

As shown in Figure 8.3, an observer at rest watching the particles go past will measure

$$\bar{\ell} = \frac{\ell}{\cosh\beta} \tag{8.3}$$

due to length contraction, and will therefore observe a number density

$$\bar{n} = n\,\cosh\beta. \tag{8.4}$$

How many such particles move past an observer at rest per unit time? The distance traveled by each particle is $v \cdot 1 = \tanh\beta$, so the *number flux* of particles is given by

$$\bar{n}\,\tanh\beta = n\,\sinh\beta. \tag{8.5}$$

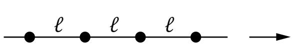


FIGURE 8.1. Equally spaced particles, moving uniformly to the right.

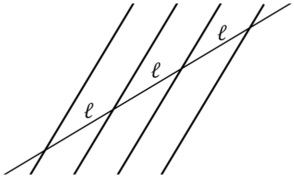


FIGURE 8.2. A spacetime diagram of equally spaced particles, moving uniformly to the right.

This leads us to define the *number flux 2-vector*

$$\boldsymbol{N} = n\,\boldsymbol{u} = \begin{pmatrix} n \cosh\beta \\ n \sinh\beta \end{pmatrix}, \tag{8.6}$$

where n is the number density of the particles *in their rest frame.*

The energy density of the particles is just their number density times the mass of each particle. Thus, in the rest frame of the particles, the energy density is just mn, whereas an observer at rest (i.e., moving with respect to the particles with $v = \tanh\beta$) would observe an energy density of

$$\bar{m}\bar{n} = mn \cosh^2\beta. \tag{8.7}$$

The energy density therefore transforms with *two* factors of $\cosh\beta$, rather than one, so that it is not the component of a vector. Rather, it is the component of a rank 2 object, the *stress-energy-momentum tensor* T, usually referred to as either the *stress tensor* or the *energy-momentum tensor.* Our previous discussion correctly suggests that the energy-momentum tensor for our collection of particles ("dust") should be given by some sort of

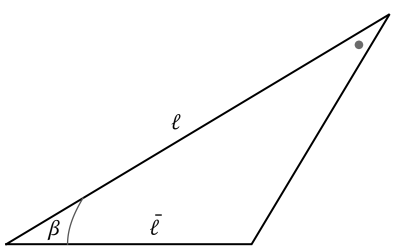


FIGURE 8.3. Determining the spacing between moving particles, as seen by an observer at rest. The dot denotes a right angle.

product of the vectors $\boldsymbol{p}$ and $\boldsymbol{N}$, and the correct answer turns out to be

$$T = \boldsymbol{p} \otimes \boldsymbol{N}, \tag{8.8}$$

where the symbol $\otimes$ means the *tensor product* and is read as "tensor." All this means is that the new object T has components T^{ij} that are given by

$$T^{ij} = p^i N^j = mn\, u^i\, u^j. \tag{8.9}$$

The matrix of components (T^{ij}) is symmetric; T is a *symmetric* (rank 2, contravariant) tensor.

We choose to avoid further discussion of the tensor nature of T by using instead a physically equivalent description in terms of vector-valued differential forms. Consider the *energy-momentum 1-forms* defined by

$$T^i = T^i{}_j\, \sigma^j, \tag{8.10}$$

where $T^i{}_j$ is (of course!) obtained from T^{ij} by lowering an index with the metric, that is,

$$T^i{}_j = g_{jk} T^{ik} = T^{ik}\, \hat{\boldsymbol{e}}_k \cdot \hat{\boldsymbol{e}}_j = \pm T^{ij}. \tag{8.11}$$

Then we can think of the energy-momentum tensor as a vector-valued 1-form, denoted by $\vec{\boldsymbol{T}}$, which is given by

$$\vec{\boldsymbol{T}} = T^i \hat{\boldsymbol{e}}_i. \tag{8.12}$$

8.2 DUST

Let's summarize the discussion in Section 8.1, but without the restriction to Minkowski space. Matter is described by the energy-momentum 1-forms

$$T^i = T^i{}_j\, \sigma^j, \tag{8.13}$$

which can be combined into a single vector-valued 1-form

$$\vec{\boldsymbol{T}} = T^i \hat{\boldsymbol{e}}_i. \tag{8.14}$$

Consider an observer with velocity vector $\vec{\boldsymbol{v}}$ satisfying

$$\vec{\boldsymbol{v}}\, d\tau = d\vec{\boldsymbol{r}} \tag{8.15}$$

and drop the arrow to write as usual

$$v = \vec{\boldsymbol{v}} \cdot d\vec{\boldsymbol{r}} \tag{8.16}$$

for the 1-form physically equivalent to $\vec{\boldsymbol{v}}$. The observer's 4-velocity $\vec{\boldsymbol{v}}$ is a future-pointing, timelike unit vector, so $\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}} = -1$ and therefore $v = -d\tau$.

Given an energy-momentum tensor $\vec{\boldsymbol{T}}$ and an observer with velocity v, it is reasonable to ask how to obtain the energy density seen by this observer. Not surprisingly, this involves some sort of dot product between $\vec{\boldsymbol{T}}$ and v, but it requires a dot product both on the vector part of $\vec{\boldsymbol{T}}$ and on its 1-form part. Begin by using the metric g to perform the latter operation, defining[1]

$$\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} = -g(\vec{\boldsymbol{T}}, v), \tag{8.17}$$

which we can write in terms of components using

$$g(\vec{\boldsymbol{T}}, v) = T^i{}_j \hat{\boldsymbol{e}}_i \, g(\sigma^j, \sigma^k) \, v_k = T^i{}_j \, v^j \, \hat{\boldsymbol{e}}_i \tag{8.18}$$

since $v = v_k \, \sigma^k$. The observed energy density is then given by

$$\rho_{\text{obs}} = -\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} \cdot \vec{\boldsymbol{v}} = g(\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}}, v), \tag{8.19}$$

where in components

$$\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}} = \left(T^i{}_j \sigma^j \hat{\boldsymbol{e}}_i\right) \cdot \left(v^k \, \hat{\boldsymbol{e}}_k\right) = T^i{}_j \sigma^j v_i \tag{8.20}$$

and where the last equality in (8.19) reflects the ability to perform these two types of dot products in either order. The condition that the energy-momentum tensor be symmetric can be expressed in the form

$$g(\vec{\boldsymbol{T}}, v) \cdot d\vec{\boldsymbol{r}} = \vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}} \tag{8.21}$$

(for all spacetime vectors $\vec{\boldsymbol{v}}$), which we will henceforth assume, and which is equivalent to $T^{ij} = T^{ji}$.

The energy-momentum tensor for *dust* with velocity vector $\vec{\boldsymbol{u}}$ is given by

$$\vec{\boldsymbol{T}} = \rho \, \vec{\boldsymbol{u}} \, u, \tag{8.22}$$

where $u = \vec{\boldsymbol{u}} \cdot d\vec{\boldsymbol{r}} = -d\lambda$, with λ denoting the proper time of the dust particles in order to avoid confusion with the proper time τ of the observer. Equivalently,

$$\vec{\boldsymbol{T}}_{\vec{\boldsymbol{u}}} = \rho \, \vec{\boldsymbol{u}}, \tag{8.23}$$

$$\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{u}} = -\rho \, u. \tag{8.24}$$

[1]The sign is chosen so that $\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}}$ is future-pointing. This definition of $\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}}$ will be used for any spacetime vector $\vec{\boldsymbol{v}}$, not necessarily timelike, although its interpretation will be different if $\vec{\boldsymbol{v}}$ is not timelike.

The *energy density* in the rest frame of the dust particles is then

$$-\vec{\boldsymbol{T}}_{\vec{\boldsymbol{u}}} \cdot \vec{\boldsymbol{u}} = \rho \tag{8.25}$$

as expected, and the observed energy density in some other frame is given by

$$-\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} \cdot \vec{\boldsymbol{v}} = \rho\,\vec{\boldsymbol{u}}\,g(u,v) \cdot \vec{\boldsymbol{v}} = \rho\,(\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}})^2 \tag{8.26}$$

since $g(u,v) = \vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}}$. We again recover the expected result, since $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = \cosh\beta$ (and $\rho = mn$ in the example considered in Section 8.1).

Dust is not very interesting, since the dust particles do not interact with each other. A slightly more general model is a *perfect fluid*, in which there is also a *pressure density* p. The energy-momentum tensor for a perfect fluid with velocity vector $\vec{\boldsymbol{u}}$ is given by

$$\vec{\boldsymbol{T}} = (\rho + p)\,\vec{\boldsymbol{u}}\,u + p\,d\vec{\boldsymbol{r}} \tag{8.27}$$

for which

$$\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{u}} = -(\rho + p)\,u + p\,u = -\rho\,u \tag{8.28}$$

as was the case for dust, so the energy density (at rest) is still ρ. In (spacelike) directions $\vec{\boldsymbol{v}}$ orthogonal to $\vec{\boldsymbol{u}}$ we have instead

$$\begin{aligned}
-\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} \cdot \vec{\boldsymbol{v}} &= g(\vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}}, v) \\
&= g(p\,v, v) \\
&= p\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{v}} \\
&= p,
\end{aligned} \tag{8.29}$$

where we have used $\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{v}} = 0$ and further assumed $|\vec{\boldsymbol{v}}| = 1$. Thus, ρ and p correspond to the energy and pressure densities[2] in the rest frame of the fluid, and p is the same in all (spatial) directions in this frame.

8.3 FIRST GUESS AT EINSTEIN'S EQUATION

In Newtonian theory, gravity is described by the gravitational potential Φ. It can be shown that tidal acceleration in the $\vec{\boldsymbol{u}}$ direction is just the

[2]The spatial components of the energy-momentum tensor can be interpreted as components of the force exerted across surfaces in various orientations; these are the *stresses* of classical mechanics. Diagonal stresses are *pressures*; off-diagonal stresses are *shears*. The mixed "space/time" components of the energy-momentum tensor are *momentum densities* and *energy fluxes*, and of course the "time/time" component is energy density. The symmetry of $\vec{\boldsymbol{T}}$ says that energy fluxes are the same as momentum densities, and that the classical stress matrix is symmetric.

difference in the acceleration due to gravity at nearby points, which can be expressed as a directional derivative of the gravitational field, namely

$$-(\vec{\boldsymbol{u}} \cdot \vec{\boldsymbol{\nabla}})(\vec{\boldsymbol{\nabla}}\Phi).$$

More precisely, this expression describes the vector change in the displacement between two freely falling objects, which are (originally) separated in the $\vec{\boldsymbol{u}}$ direction; this is geodesic deviation. In relativity, geodesic deviation is described by an expression of the form

$$-R^i{}_{mjn} v^m u^j v^n.$$

This suggests that there is a correspondence between second derivatives of Φ (in the i and j directions) and $R^i{}_{mjn} v^m v^n$.

As we have seen, energy density in relativity is described by

$$\rho = g(\vec{\boldsymbol{T}}, \vec{\boldsymbol{v}} \cdot d\vec{\boldsymbol{r}}) \cdot \vec{\boldsymbol{v}} = T^i{}_j \, v_i \, v^j, \tag{8.30}$$

whereas in Newtonian theory matter enters through *Poisson's equation*

$$\Delta\Phi = 4\pi\rho, \tag{8.31}$$

where ρ is the matter density (energy density). The Laplacian can be thought of as the trace of a matrix of second derivatives. If second derivatives of Φ correspond to the Riemann tensor, then the Laplacian corresponds to the trace of the Riemann tensor, which is known as the Ricci tensor, whose components are given by

$$R_{ij} = R^m{}_{imj}. \tag{8.32}$$

Thus, a reasonable guess for Einstein's equation, the relationship between curvature and matter, would be

$$R_{ij} = 4\pi T_{ij} = \frac{4\pi G}{c^2} T_{ij}. \tag{8.33}$$

This guess turns out to be correct only in the absence of matter, when $\rho = 0$; Einstein's equation in vacuum is indeed (equivalent to)

$$R_{ij} = 0. \tag{8.34}$$

However, (8.33) turns out to be awkward mathematically and incorrect physically. The mathematical objection is that the Ricci tensor is naturally a (rank 2, covariant) tensor, while the energy-momentum tensor is naturally

a vector-valued 1-form. We are accustomed to raising and lowering indices by now, so this objection is not necessarily serious, but it turns out that we can do better.

The physical objection is far more serious: Conservation of matter (discussed in Section 8.4) together with the contracted Bianchi identity (Section 18.7) would force the *Ricci scalar*

$$R = R^m{}_m \tag{8.35}$$

to be constant. It is unreasonable to expect that highly nonsymmetric configurations of matter would correspond to spacetime geometries with constant (scalar) curvature.

8.4 CONSERVATION LAWS

We have seen that the energy-momentum of a cloud of dust, according to an observer with 4-velocity $\vec{\boldsymbol{v}}$, is given by[3]

$$\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} = \rho\,\vec{\boldsymbol{u}}\,(-\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}}). \tag{8.36}$$

If the dust is moving at speed $v = \tanh\beta$ with respect to the observer $\vec{\boldsymbol{v}}$, then the 4-velocity $\vec{\boldsymbol{u}}$ of the dust takes the form

$$\vec{\boldsymbol{u}} = \begin{pmatrix} 1 \\ \bar{\boldsymbol{u}} \end{pmatrix} \cosh\beta, \tag{8.37}$$

where we have introduced the notation $\bar{\boldsymbol{u}}$ for the *3-velocity* of the dust. In this frame, $\vec{\boldsymbol{v}}$ is the time direction, so

$$\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} = -\cosh\beta \tag{8.38}$$

and we have

$$\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} = \rho \begin{pmatrix} 1 \\ \bar{\boldsymbol{u}} \end{pmatrix} \cosh^2\beta = \rho_{\text{obs}} \begin{pmatrix} 1 \\ \bar{\boldsymbol{u}} \end{pmatrix}. \tag{8.39}$$

We expect a classical conservation law along the lines of

$$\dot{\rho}_{\text{obs}} + \bar{\boldsymbol{\nabla}}\cdot(\rho_{\text{obs}}\,\bar{\boldsymbol{u}}) = 0, \tag{8.40}$$

where $\bar{\boldsymbol{\nabla}}$ denotes the *three-dimensional* gradient, which says that the matter density can change only if the matter flow "diverges." This conservation law is essentially the requirement that

$$\vec{\boldsymbol{\nabla}}\cdot\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} = 0, \tag{8.41}$$

[3] Note that $-\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{v}} > 0$ if $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ are future-pointing, timelike vectors.

where this divergence operator is four-dimensional (and therefore includes a t-derivative). But

$$\vec{\boldsymbol{T}}_{\vec{\boldsymbol{v}}} = -T^i{}_j \, v^j \, \hat{\boldsymbol{e}}_i \tag{8.42}$$

and conservation should hold for *any* $\vec{\boldsymbol{v}}$. It therefore seems reasonable to expect a conservation condition of the form

$$\vec{\boldsymbol{\nabla}} \cdot (T^i{}_j \, \hat{\boldsymbol{e}}_i) = 0. \tag{8.43}$$

This is a divergence on the *vector* part of $\vec{\boldsymbol{T}}$, but the symmetry of the energy-momentum tensor allows us to regard this as a divergence on the *1-form* part of $\vec{\boldsymbol{T}}$ instead, that is, to expect a conservation condition of the form

$$*d{*}T^i = *d{*}(T^i{}_j \, \sigma^j) = 0. \tag{8.44}$$

The equivalence between (8.43) and (8.44) in special relativity (and Minkowski coordinates) can be easily checked, since both notions of divergence involve simply taking the x^i-derivative of the ith component; the symmetry of the energy momentum tensor allows us to perform this sum on either the first (vector) or second (1-form) "slot" of $T^i{}_j$. In curved space (and/or curvilinear coordinates), the above argument fails. However, there is an obvious alternative, which turns out to be correct: Demand conservation in the form

$$*d{*}\vec{\boldsymbol{T}} = \vec{\boldsymbol{0}}, \tag{8.45}$$

where d now acts on both the 1-form and the vector parts of $\vec{\boldsymbol{T}}$.

It is straightforward to run this argument backward and recover the expected conservation law (8.40) for dust in Minkowski space. First of all, in Minkowski coordinates, the basis vectors are constant, so that (8.45) implies

$$*d{*}T^i = 0 \tag{8.46}$$

for each energy-momentum 1-form

$$T^i = \rho \, u^i \, u_j \, \sigma^j. \tag{8.47}$$

But the components of $\vec{\boldsymbol{u}}$ are given by (8.37), so that

$$T^t = \rho_{\text{obs}} \, (-dt + \bar{u}_i \, dx^i), \tag{8.48}$$

$$T^j = \bar{u}^j T^t \tag{8.49}$$

where i, j now run only over the three spatial indices. Setting

$$*d{*}T^t = 0 \tag{8.50}$$

leads immediately to (8.40), which can be thought of as conservation of energy. Inserting this result into

$$*d{*}T^j = 0 \tag{8.51}$$

leads to

$$\rho_{\text{obs}} \left(\dot{\bar{u}}^i + \bar{u}^j \, \frac{\partial \bar{u}^i}{\partial x^j} \right) = 0 \tag{8.52}$$

or equivalently

$$\dot{\bar{\boldsymbol{u}}} + (\bar{\boldsymbol{u}} \cdot \bar{\boldsymbol{\nabla}})\bar{\boldsymbol{u}} = \bar{\mathbf{0}}, \tag{8.53}$$

which is the well-known *Navier–Stokes equation*, and can be thought of as a statement of conservation of momentum.

We conclude that (8.45) does indeed provide a reasonable notion of conservation.

8.5 THE EINSTEIN TENSOR

As discussed in Sections 8.1–8.4, *matter* in general relativity is described by a vector-valued 1-form

$$\vec{\boldsymbol{T}} = T^i \, \hat{\boldsymbol{e}}_i, \tag{8.54}$$

where

$$T^i = T^i{}_j \, \sigma^j \tag{8.55}$$

are the *energy-momentum 1-forms*. The components $T^i{}_j$ are (also) components of the *energy-momentum tensor*. We assume that the energy-momentum tensor is *symmetric*, that is, that

$$T_{ji} = T_{ij}, \tag{8.56}$$

which can also be expressed as the requirement that

$$g(\vec{\boldsymbol{T}}, v) \cdot d\vec{\boldsymbol{r}} = \vec{\boldsymbol{T}} \cdot \vec{\boldsymbol{v}} \tag{8.57}$$

for *any* vector $\vec{\boldsymbol{v}}$. We further assume that matter is conserved, by which we mean that the energy-momentum tensor is *divergence free*, that is, that

$$*d{*}\vec{\boldsymbol{T}} = \vec{\mathbf{0}}. \tag{8.58}$$

Introducing the *energy-momentum 3-forms*

$$\tau^i = {*}T^i \tag{8.59}$$

brings the conservation condition to the form

$$d(\tau^i\,\hat{\boldsymbol{e}}_i) = 0. \tag{8.60}$$

Finally, expanding $d\hat{\boldsymbol{e}}_i$ in terms of the connection 1-forms and equating components leads to

$$d\tau^i + \omega^i{}_j \wedge \tau^j = 0 \tag{8.61}$$

as the condition for matter to be conserved.

We would like to relate matter to curvature and therefore ask whether we can construct a "curvature 3-form" that satisfies (8.61). As discussed in Section 18.7, the curvature 2-forms satisfy the (second) Bianchi identity, namely,

$$d\Omega^i{}_j + \omega^i{}_k \wedge \Omega^k{}_j - \Omega^i{}_k \wedge \omega^k{}_j = 0. \tag{8.62}$$

The simplest 3-form we can construct from the curvature 2-forms is

$$\alpha^i = \Omega^i{}_j \wedge \sigma^j \tag{8.63}$$

and we can now compute

$$\begin{aligned} d\alpha^i &= d\Omega^i{}_j \wedge \sigma^j + \Omega^i{}_j \wedge d\sigma^j \\ &= (-\omega^i{}_k \wedge \Omega^k{}_j + \Omega^i{}_k \wedge \omega^k{}_j) \wedge \sigma^j - \Omega^i{}_j \wedge \omega^j{}_n \wedge \sigma^n, \end{aligned} \tag{8.64}$$

where we have used the first structure equation to differentiate σ^j. The last two terms cancel, and we are indeed left with (8.61), but for α^i rather than τ^i. It appears that we have found our "curvature 3-form!"! However, the (first) Bianchi identity (18.52) says precisely that $\alpha^i = 0$, so this is not in fact what we were looking for.

Nonetheless, we're almost there. What is needed is a slightly different combination of terms of the same form, that is, which look like $\Omega^i{}_j \wedge \sigma^m$. The correct choice turns out to be the *Einstein 3-forms*

$$\gamma^i = -\frac{1}{2}\Omega_{jk} \wedge *(\sigma^i \wedge \sigma^j \wedge \sigma^k), \tag{8.65}$$

which indeed satisfy

$$d\gamma^i + \omega^i{}_j \wedge \gamma^j = 0. \tag{8.66}$$

To verify (8.66) requires writing out all of the terms on the left-hand side of this expression, using the fact that the indices i, j, k must be distinct in the definition of γ^i, (8.65), and showing that the terms cancel in pairs; see Section A.6 for further details.

By analogy with energy-momentum, we can work backward and construct the *Einstein 1-forms*

$$G^i = *\gamma^i = G^i{}_j\,\sigma^j \tag{8.67}$$

whose components $G^i{}_j$ define the *Einstein tensor*, which we can think of as a vector-valued 1-form

$$\vec{\boldsymbol{G}} = G^i\hat{\boldsymbol{e}}_i = G^i{}_j\,\sigma^j\hat{\boldsymbol{e}}_i. \tag{8.68}$$

The Einstein tensor is closely related to the Ricci tensor, and in fact

$$G^i{}_j = R^i{}_j - \frac{1}{2}\delta^i{}_j\,R, \tag{8.69}$$

where $R = R^m{}_m$ is the Ricci scalar; see Section A.5 for further details. This equation is more commonly written in covariant form, that is, with all indices down, namely

$$G_{ij} = R_{ij} - \frac{1}{2}g_{ij}\,R. \tag{8.70}$$

The Einstein tensor is often referred to as the "trace-reversed" Ricci tensor, since

$$G^m{}_m = R^m{}_m - \frac{1}{2}4R = -R^m{}_m. \tag{8.71}$$

8.6 EINSTEIN'S EQUATION

Summarizing the discussion so far, matter is described by the energy-momentum tensor, which we now write as

$$\vec{\boldsymbol{T}} = T^i{}_j\sigma^j\hat{\boldsymbol{e}}_i = *\tau^i\hat{\boldsymbol{e}}_i, \tag{8.72}$$

and conservation of matter requires that $\vec{\boldsymbol{T}}$ be divergence-free, that is, that

$$d{*}\vec{\boldsymbol{T}} = \vec{\boldsymbol{0}}. \tag{8.73}$$

Meanwhile, it turns out that there is a *unique* divergence-free vector-valued 1-form that can be constructed from the curvature, namely the Einstein tensor

$$\vec{\boldsymbol{G}} = G^i{}_j\sigma^j\hat{\boldsymbol{e}}_i = *\gamma^i\hat{\boldsymbol{e}}_i, \tag{8.74}$$

which satisfies

$$d{*}\vec{\boldsymbol{G}} = \vec{\boldsymbol{0}}. \tag{8.75}$$

Einstein's fundamental insight was to equate these two expressions; *Einstein's equation* is

$$\vec{\boldsymbol{G}} = 8\pi\vec{\boldsymbol{T}} = \frac{8\pi G}{c^2}\vec{\boldsymbol{T}}, \tag{8.76}$$

where the factor of 8π is chosen to ensure agreement with Newtonian theory in the appropriate limit. In more traditional tensor language, Einstein's equation is written as

$$G_{ij} = R_{ij} - \frac{1}{2}g_{ij}R = \frac{8\pi G}{c^2}T_{ij}. \tag{8.77}$$

Curvature equals matter! The conservation of matter has become a consequence of the Bianchi identity!

8.7 THE COSMOLOGICAL CONSTANT

Einstein was not happy with some of the predictions of his theory, notably, as we will see in Section 9.5, that the universe is expanding. He therefore looked for a way to modify his theory to permit a static universe.

There is in fact just one way to do so; the only other divergence-free geometric object (of the correct rank, that is, with the correct number of components) is the line element itself! Using the fact that

$$*\sigma^i = \pm\sigma^j \wedge \sigma^k \wedge \sigma^\ell \tag{8.78}$$

for an appropriate choice of j, k, ℓ, direct computation shows that

$$d{*}d\vec{\boldsymbol{r}} = \vec{\boldsymbol{0}}, \tag{8.79}$$

as shown in Section A.8. Thus, a generalization of Einstein's equation is

$$\vec{\boldsymbol{G}} + \Lambda\, d\vec{\boldsymbol{r}} = 8\pi\vec{\boldsymbol{T}}, \tag{8.80}$$

where Λ is the *cosmological constant*. In more traditional tensor language, Einstein's equation with cosmological constant is written as

$$G_{ij} + \Lambda\, g_{ij} = \frac{8\pi G}{c^2}T_{ij}. \tag{8.81}$$

As we will discuss in Section 9.5, suitable values of Λ do indeed permit a static universe. After astronomers later observed that the universe is in fact expanding, Einstein referred to the cosmological constant as the biggest mistake of his career. However, recent observations suggest that the cosmological constant is small but nonzero; perhaps Einstein deserves the last laugh after all.

8.8 PROBLEMS

1. **Gravitational Waves**

   Consider the line element

$$ds^2 = -du\,dv + f(u)^2\,dy^2 + g(u)^2\,dz^2.$$

   (a) Find the curvature of this geometry, that is, determine the (nonzero, independent) curvature 2-forms and/or the components of the Riemann curvature tensor.

   Can you find an orthonormal basis of 1-forms?

   (b) Determine the (independent, nonzero) components of Einstein's vacuum equation (without cosmological constant) for this geometry.

   (c) Find an explicit, nontrivial solution of the equation just obtained, that is, find explicit functions $f(u)$ and $g(u)$ so that the above geometry is a vacuum solution of Einstein's equation.

The trivial solution is Minkowski space. Nontrivial solutions model (infinite) gravitational waves. Which way are they traveling? How fast? You should be able to see the quadrupole nature of gravitational waves by examining the signs of your nonzero curvature terms, which should suggest "shrinking" in one direction while "stretching" in the other.

≻ Chapter 9 ≺

Cosmological Models

9.1 Cosmology

Cosmology is, simply put, the study of the universe. Relativistic cosmology is the study of solutions of Einstein's equation, which may represent the broad features of the universe as a whole.

One constraint on such models is the seemingly innocuous statement that the sky is dark at night. Why is this surprising? In a Euclidean universe with a uniform distribution of stars, both the density of matter and the apparent luminosity fall off as $1/r^2$. Thus, the perceived brightness of every thin spherical shell of stars, as seen at the center of the sphere, is the same; an infinite universe would therefore lead to an infinitely bright night sky! This seeming contradiction is known as *Olbers' paradox*.

How can we avoid Olbers' paradox? Perhaps the universe is finite. Or perhaps the spatial distribution of stars is not uniform. Or perhaps the universe is not Euclidean. Or perhaps the universe is not constant in time.

Any reasonable theory of cosmology must provide a resolution to this conundrum.

9.2 The Cosmological Principle

The simplest cosmological models are based on the principle that the universe is the same everywhere. This principle is sometimes referred to as the *cosmological principle* and can be thought of as an extreme generalization of the Copernican view that the Earth is not the center of the universe.

More precisely, we assume that the universe is *homogeneous* and *isotropic*. Homogeneity is the assumption that there exists a foliation of spacetime into spacelike surfaces ("instants of cosmic time"), each of which possesses no privileged points. Isotropy is the assumption that there exists a foliation of spacetime into timelike curves ("cosmic observers"), each of which sees no privileged directions at any time.

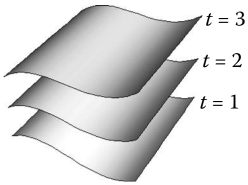


FIGURE 9.1. A foliation of spacetime by hypersurfaces of constant cosmic time.

More formally, a spacetime is *(spatially) homogeneous* if it can be foliated by spacelike hypersurfaces Σ_t, as illustrated in Figure 9.1, such that all points in a given surface Σ_t are equivalent. Thus, there is no preferred point in Σ_t; physics is therefore the same everywhere at each instant of "time." There is, however, no requirement that physics at different times also be the same.

Similarly, a spacetime is *(spatially) isotropic* at each point if there exists a family of observers $\vec{\boldsymbol{u}}_s$, as illustrated in Figure 9.2, such that all spatial directions (that is, all directions perpendicular to $\vec{\boldsymbol{u}}_s$) are equivalent. Thus, at every point, there is no preferred (spatial) direction.

The twin requirements of homogeneity and isotropy imply that the worldlines of the cosmic observers are orthogonal to the cosmic surfaces, as illustrated in Figure 9.3; if not, then the projection of $\vec{\boldsymbol{u}}_s$ into Σ_t would be nonzero and would define a preferred direction. Thus, "cosmic time" is precisely the time measured by such observers. This condition is often assumed separately and is known as *Weyl's postulate*.

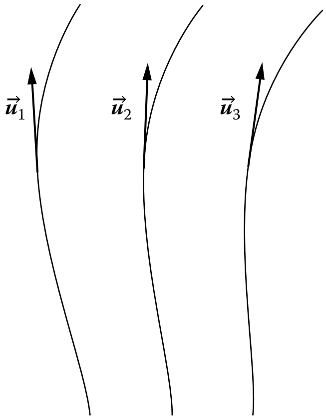


FIGURE 9.2. A family of cosmic observers.

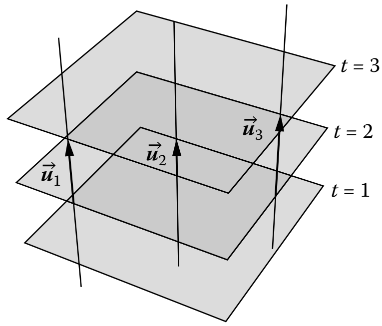


FIGURE 9.3. In a spacetime that is both isotropic and homogeneous, the cosmic observers are orthogonal to the cosmic surfaces.

9.3 CONSTANT CURVATURE

In any dimension and signature, it turns out there is a unique (local) geometry of constant (sectional) curvature, which can be classified by the value of the scalar curvature R. The three possibilities are illustrated for two dimensions with Euclidean signature in Figure 9.4, showing a plane (zero curvature), sphere (constant positive curvature), and Lorentzian hyperboloid (constant negative curvature), respectively. In each case, the *surface* is two-dimensional, even though the representation shown is embedded in three dimensions. Furthermore, the hyperboloid is a Riemannian surface (its metric is positive-definite), even though it is shown embedded in three-dimensional Minkowski space. An isometric representation does exist in three-dimensional Euclidean space, namely the *pseudosphere*,[1] but the analogy to the sphere is less apparent in this representation.

We consider here the case of *three*-dimensional surfaces with Euclidean signature, which can also be represented as in Figure 9.4, but with one dimension suppressed.

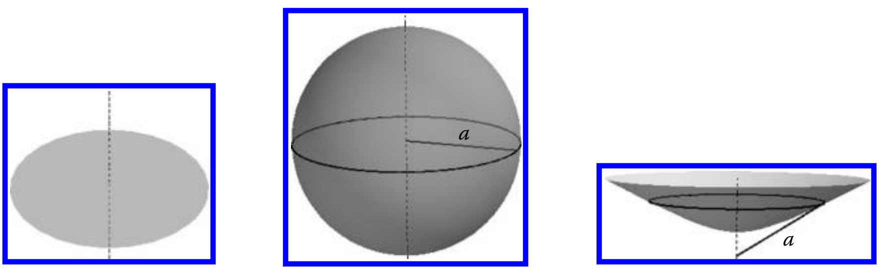


FIGURE 9.4. The two-dimensional surfaces of constant curvature.

[1]See, for example, Chapter 14 of [2] or Figure 18.13.

ZERO CURVATURE

If $R = 0$, then the geometry is *flat*; this is ordinary Euclidean space. The line element can be brought to the form

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 \\ &= dr^2 + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right) \\ &= d\psi^2 + \psi^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right), \end{aligned} \tag{9.1}$$

where $r = \psi$. This geometry is just (Euclidean) $\mathbb{R}^3$.

POSITIVE CURVATURE

If $R = 6/a^2 > 0$, then the geometry is *spherical*. The line element can be brought to the form

$$\begin{aligned} ds^2 &= a^2 \left(d\psi^2 + \sin^2\psi \left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right) \\ &= a^2 \left(\frac{dr^2}{1-r^2} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right), \end{aligned} \tag{9.2}$$

where $r = \sin\psi$. We can embed this surface in Euclidean $\mathbb{R}^4$ with coordinates (w,x,y,z) as the set of points satisfying

$$w^2 + x^2 + y^2 + z^2 = a^2. \tag{9.3}$$

NEGATIVE CURVATURE

If $R = -6/a^2 < 0$, then the geometry is *hyperbolic*. The line element can be brought to the form

$$\begin{aligned} ds^2 &= a^2 \left(d\psi^2 + \sinh^2\psi \left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right) \\ &= a^2 \left(\frac{dr^2}{1+r^2} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right), \end{aligned} \tag{9.4}$$

where $r = \sinh\psi$. We can embed this surface in Minkowskian $\mathbb{R}^4$ with coordinates (T,x,y,z) as the set of points satisfying

$$T^2 - x^2 - y^2 - z^2 = a^2. \tag{9.5}$$

All three cases can be written simultaneously as

$$ds^2 = a^2 \left(\frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right), \tag{9.6}$$

where $k = a^2R/6 \in \{-1, 0, 1\}$. (The value of a corresponds to an irrelevant scale if $R = 0$.)

9.4 ROBERTSON–WALKER METRICS

We now study simple models of the universe and therefore assume both homogeneity and isotropy. As discussed in Section 9.2, homogeneity implies that spacetime is foliated by spacelike hypersurfaces Σ_t; isotropy implies there are preferred "cosmic observers" orthogonal to these hypersurfaces. Thus, each Σ_t represents an instant of time according to these cosmic observers. We can therefore assume without loss of generality that the surfaces are labeled using "cosmic time," that is, that t is proper time according to these cosmic observers.

The line element therefore takes the form

$$ds^2 = -dt^2 + h_{ij}\, dx^i\, dx^j, \tag{9.7}$$

where h_{ij} are the components of the line element restricted to Σ_t. There are no cross terms in (9.7), since the worldlines of cosmic observers are orthogonal to Σ_t.

Homogeneity implies that each Σ_t has no privileged points; among other things, this means that the curvature must be constant. Thus, each Σ_t must be a surface of constant curvature, as discussed in Section 9.3. The line element therefore becomes

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2 \right) \right), \tag{9.8}$$

which is known as the *Robertson–Walker metric*. The free parameters are k, which determines the shape of space, and $a(t)$, which determines the scale of the universe as a function of cosmic time. You can think of the universe as the surface of a balloon, whose "radius" is given by $a(t)$, although this analogy is only strictly true for the spherical case ($k = 1$), which is illustrated in Figure 9.5. The parameter k also determines the *topology* of the universe; the universe is *closed* (finite) if $k = 1$, and *open* (infinite) otherwise.

It is straightforward to compute the Einstein tensor for the Robertson–Walker metric. As shown in Section A.9, the independent, nonzero curvature 2-forms are

$$\Omega^t{}_r = \frac{\ddot{a}}{a}\sigma^t \wedge \sigma^r, \tag{9.9}$$

$$\Omega^t{}_\theta = \frac{\ddot{a}}{a}\sigma^t \wedge \sigma^\theta, \tag{9.10}$$

$$\Omega^t{}_\phi = \frac{\ddot{a}}{a}\sigma^t \wedge \sigma^\theta, \tag{9.11}$$

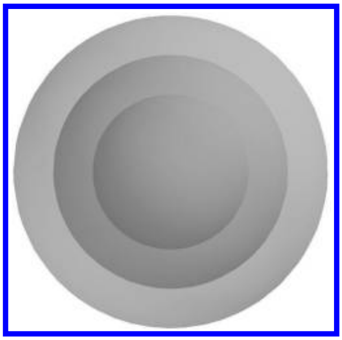


FIGURE 9.5. An expanding spherical balloon, shown at 3 different instants of time.

$$\Omega^r{}_\theta = \frac{\dot{a}^2 + k}{a^2}\sigma^r \wedge \sigma^\theta, \tag{9.12}$$

$$\Omega^r{}_\phi = \frac{\dot{a}^2 + k}{a^2}\sigma^r \wedge \sigma^\phi, \tag{9.13}$$

$$\Omega^\theta{}_\phi = \frac{\dot{a}^2 + k}{a^2}\sigma^\theta \wedge \sigma^\phi, \tag{9.14}$$

and the nonzero components of the Einstein tensor are

$$G^t{}_t = -3\left(\frac{\dot{a}^2 + k}{a^2}\right), \tag{9.15}$$

$$G^r{}_r = G^\theta{}_\theta = G^\phi{}_\phi = -\left(\frac{2a\ddot{a} + \dot{a}^2 + k}{a^2}\right). \tag{9.16}$$

Thus, the Einstein tensor has precisely the form of the energy-momentum tensor of a perfect fluid! This shouldn't be a surprise, since we have assumed homogeneity and isotropy. Inserting the Einstein tensor into Einstein's equation (with cosmological constant), and assuming the energy-momentum tensor is indeed a perfect fluid, we obtain

$$-3\left(\frac{\dot{a}^2 + k}{a^2}\right) + \Lambda = -8\pi\rho, \tag{9.17}$$

$$-\left(\frac{2a\ddot{a} + \dot{a}^2 + k}{a^2}\right) + \Lambda = 8\pi p, \tag{9.18}$$

where ρ is the energy density and p the pressure density, and where both of these densities are functions of time t.

9.5 THE BIG BANG

Slightly rewriting Einstein's equation for the Robertson–Walker metric, as given in Section 9.4, we obtain first

$$\frac{\dot{a}^2 + k}{a^2} = \frac{8\pi\rho + \Lambda}{3} \tag{9.19}$$

and then

$$\frac{2\ddot{a}}{a} = \Lambda - 8\pi p - \frac{\dot{a}^2 + k}{a^2} = \frac{2}{3}\Lambda - \frac{8\pi}{3}\left(\rho + 3p\right). \tag{9.20}$$

A physically realistic model that is not empty will have strictly positive energy density ($\rho > 0$) and nonnegative pressure density ($p \geq 0$).

Einstein initially assumed that $\Lambda = 0$. In this case (and also if $\Lambda < 0$), we must have

$$\ddot{a} < 0. \tag{9.21}$$

In other words, in the absence of a cosmological constant, the universe cannot be static! Einstein was unhappy with this conclusion, since the then-current belief was that the universe was indeed static; he later added the cosmological constant term precisely so as to make static solutions possible.

However, later observations showed that the universe is currently expanding, that is, that

$$\dot{a}\Big|_{\text{now}} > 0. \tag{9.22}$$

In an expanding universe, we expect to see all other stellar objects receding from us. To see that this property does not violate our assumptions of homogeneity and isotropy, consider again an expanding balloon, shown at two different stages in Figure 9.6. Imagine that the surface of the balloon represents our universe, and the black dots represent galaxies. As the balloon expands, the distance between *every* pair of galaxies increases.

Elementary calculus shows that if $\dot{a} > 0$ now, and if $\ddot{a} < 0$ always, then it must be true that a was zero at some time in the past. Thus, Robertson–Walker models with $\Lambda \leq 0$ (and with expansion now) must have a past singularity, called the *Big Bang*, when the universe had zero size.

This result is a special case of a much more general principle in relativity, which, loosely stated, says that "gravity attracts." More precisely, reasonable matter (positive energy density) *always* leads to a singularity somewhere, either in the past (e.g., the Big Bang) or in the future (such as a "Big Crunch" for cosmological models, or a black hole formed by a collapsing star).

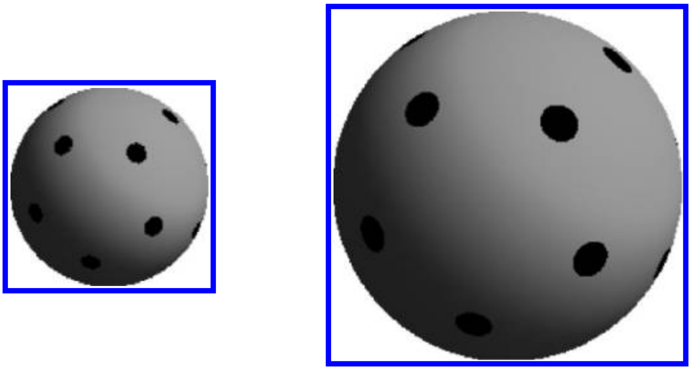


FIGURE 9.6. Distances between stellar objects increase as the universe expands, illustrated as the distance between dots on an expanding balloon.

9.6 FRIEDMANN MODELS

Current observations show that

$$\rho \gg p \tag{9.23}$$

(in appropriate units!); the universe is presently "matter dominated." We therefore consider models with

$$p = 0, \tag{9.24}$$

which are called *Friedmann–Robertson–Walker models*, or just *Friedmann models*.

With $p = 0$, the spatial component of Einstein's equation becomes

$$2a\ddot{a} + \dot{a}^2 + k = \Lambda a^2. \tag{9.25}$$

Remarkably, this equation can be integrated exactly, which is most easily done after first multiplying both sides by $\dot{a}$. After integration, we obtain

$$a\dot{a}^2 + ka = \frac{\Lambda a^3}{3} + C \tag{9.26}$$

where C is an integration constant. Rewriting this equation slightly yields *Friedmann's equation*, namely

$$\dot{a}^2 = \frac{C}{a} + \frac{\Lambda a^2}{3} - k. \tag{9.27}$$

But the time component of Einstein's equation tells us that

$$a(\dot{a}^2 + k) = \frac{8\pi\rho + \Lambda}{3}\, a^3 \tag{9.28}$$

and comparison with (9.26) implies that

$$C = \frac{8\pi\rho a^3}{3}. \tag{9.29}$$

Thus, the energy density ρ goes like $1/a^3$, which seems reasonable, since a^3 has the dimensions of volume.

9.7 Friedmann Vacuum Cosmologies

We first consider Robertson–Walker cosmologies that are also *vacuum solutions*, that is, for which[2]

$$\rho = 0 = p. \tag{9.30}$$

Friedmann's equation reduces to

$$\dot{a}^2 = \frac{\Lambda a^2}{3} - k. \tag{9.31}$$

If $k = 0$, then either $\Lambda = 0$, in which case a is constant and we have Minkowski space, or $\Lambda > 0$, in which case

$$a = e^{\pm qt} \tag{9.32}$$

where

$$q = \sqrt{\frac{|\Lambda|}{3}}. \tag{9.33}$$

We choose the exponent to be positive, so that the universe is *expanding*, to match current observations; the resulting spacetime is called *de Sitter space*.

Similarly, if $\Lambda = 0$, then (9.31) implies either that $k = 0$, which we have already considered, or that $k = -1$ and

$$a = \pm t. \tag{9.34}$$

However, the resulting spacetime turns out to be Minkowski space in disguise, as can be seen by making the coordinate transformation

$$R = t \sinh\psi = tr, \tag{9.35}$$

$$T = t \cosh\psi = t\sqrt{1 + r^2} \tag{9.36}$$

[2]Such solutions are automatically Friedmann models, since $p = 0$.

	$\mathbf{\Lambda < 0}$	$\mathbf{\Lambda = 0}$	$\mathbf{\Lambda > 0}$
$\boldsymbol{k = 1}$	(no solution)	(no solution)	de Sitter $(a = \cosh(qt)/q)$
$\boldsymbol{k = 0}$	(no solution)	Minkowski $(a = 1)$	de Sitter $(a = \exp(qt))$
$\boldsymbol{k = -1}$	anti de Sitter $(a = \sin(qt)/q)$	Minkowski $(a = t)$	de Sitter $(a = \sinh(qt)/q)$

TABLE 9.7. Classification of vacuum Friedmann models.

after which the line element becomes

$$ds^2 = -dT^2 + dR^2 + R^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right). \tag{9.37}$$

The remaining cases can be handled similarly; the result is shown in Table 9.7. Remarkably, even though there are in principle nine possible combinations of k and Λ, there are only three vacuum Friedmann solutions; different entries in the table for the same geometry correspond to different ways of slicing up that geometry.

9.8 MISSING MATTER

Non-vacuum Friedmann solutions can be classified using similar methods to those used in the vacuum case. We restrict the discussion here to some special cases. In particular, we assume throughout this section that $\Lambda = 0$, which brings Friedmann's equation to the form

$$\dot{a}^2 = \frac{8\pi\rho a^2}{3} - k. \tag{9.38}$$

Solving for k results in

$$k = \frac{8\pi\rho a^2}{3} - \dot{a}^2. \tag{9.39}$$

If we define the *critical density* by

$$\rho_c = \frac{3}{8\pi}\frac{\dot{a}^2}{a^2}\bigg|_{\text{now}}, \tag{9.40}$$

then

$$k = \frac{8\pi a^2}{3}\left(\rho - \rho_c\right). \tag{9.41}$$

Thus, in these models, the shape of the universe can be determined by comparing the observed value ρ of the energy density to the critical value ρ_c: If $\rho > \rho_c$, then $k = 1$; if $\rho = \rho_c$, then $k = 0$; and if $\rho < \rho_c$, then $k = -1$.

As discussed in Section 9.10, the value of $\frac{\dot{a}}{a}$, and hence of ρ_c, is related to the cosmological redshift; the current value is approximately

$$\rho_c \approx 2 \times 10^{-29} \frac{\text{g}}{\text{cm}}. \tag{9.42}$$

Meanwhile, the estimated average energy density of the galaxies we can see is approximately

$$\rho_{\text{galaxies}} \approx 1 \times 10^{-30} \frac{\text{g}}{\text{cm}}. \tag{9.43}$$

These numbers are astonishingly close! It is currently an open question whether there is enough matter in the universe in other forms to lead to a closed universe ($k = 1$), a mystery commonly referred to as the problem of "missing matter."

9.9 THE STANDARD MODELS

The *standard cosmological models* are models without cosmological constant ($\Lambda = 0$). We first consider Friedmann models ($p = 0$), for which

$$\dot{a}^2 = \frac{C}{a} - k. \tag{9.44}$$

If $k = 0$, we have

$$\sqrt{a}\dot{a} = \sqrt{C} \tag{9.45}$$

so that

$$a = \left(\frac{9Ct^2}{4}\right)^{1/3}, \tag{9.46}$$

which is plotted in Figure 9.8. This model is called the *Einstein–de Sitter cosmology*, and is flat, but nonetheless expanding.

If instead $k = 1$, we have

$$\dot{a}^2 = \frac{C}{a} - 1, \tag{9.47}$$

which can be solved parametrically, yielding

$$\frac{t}{C} = \frac{1}{2}(\eta - \sin\eta), \tag{9.48}$$

$$\frac{a}{C} = \frac{1}{2}(1 - \cos\eta), \tag{9.49}$$

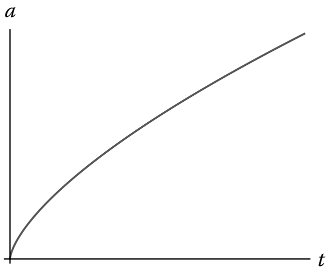


FIGURE 9.8. The expansion of the Einstein–de Sitter cosmology.

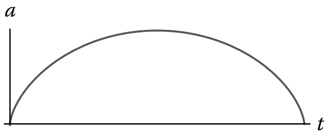


FIGURE 9.9. The expansion of the dust-filled Robertson–Walker cosmology.

which is plotted in Figure 9.9. This model is called the *dust-filled Robertson–Walker cosmology*, and is closed; this universe ends with a "Big Crunch."

If $k = -1$, the solution is formally similar, but with circular trigonometric functions replaced by hyperbolic trigonometric functions. The parametric solution in this case is

$$\frac{t}{C} = \frac{1}{2}\left(\sinh\eta - \eta\right), \tag{9.50}$$

$$\frac{a}{C} = \frac{1}{2}\left(\cosh\eta - 1\right), \tag{9.51}$$

which is plotted in Figure 9.10. This model is open (expands forever).

Finally, if $k = 1$ and

$$p = \frac{\rho}{3}, \tag{9.52}$$

we have the *radiation-filled Robertson–Walker cosmology*, for which it turns out that

$$\rho a^4 = \text{constant} \tag{9.53}$$

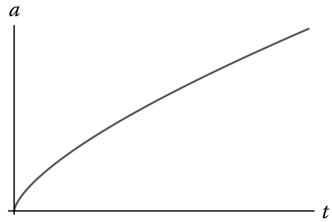


FIGURE 9.10. The hyperbolic analog of the dust-filled Robertson–Walker cosmology.

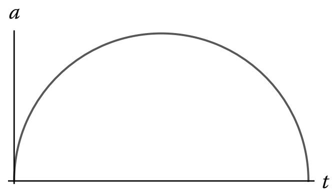


FIGURE 9.11. The expansion of the radiation-filled Robertson–Walker cosmology.

and which admits a parametric solution of the form

$$\frac{t}{B} = 1 - \cos\eta, \tag{9.54}$$

$$\frac{a}{B} = \sin\eta, \tag{9.55}$$

which is plotted in Figure 9.11, and which is again closed.

9.10 COSMOLOGICAL REDSHIFT

In special relativity, light emitted by a moving observer at one frequency is received by a stationary observer at another frequency. This is the *Doppler effect*, which can be computed using Figure 9.12. In this case, the two heavy lines correspond to observers moving at different velocities, and the two lighter lines correspond to pulses of light, emitted by the moving observer (on the right) and received by the stationary observer (on the left). Simple (hyperbolic) triangle trigonometry can be used to compute the ratio of the time intervals between the two pulses as seen by the two observers.

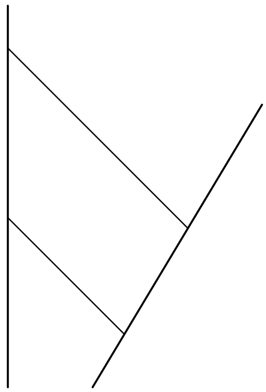


FIGURE 9.12. Spacetime diagram for computing redshift.

The same diagram can be used to compute the cosmological redshift, but the interpretation is quite different. The two heavy lines now correspond to two observers in the *same* reference frame; since the universe is expanding, these observers are *not* at rest with respect to each other. Each of these worldlines can therefore be described by $r =$ constant.

Along each light beam, the line element (9.8) takes the form

$$0 = ds^2 = -dt^2 + \frac{a^2\,dr^2}{1-kr^2} = -dt^2 + a^2\,d\psi^2 \tag{9.56}$$

so that

$$d\psi = -\frac{dt}{a}, \tag{9.57}$$

where the sign is chosen to imply an *incoming* light beam from a galaxy far away. This expression can be integrated along each light beam, but the answer must be the same in both cases since $\psi =$ constant along each worldline. If the light beams are emitted close together, the rescaled time differences dt/a between pulses must therefore be the same for both observers, which we write as

$$\frac{dt_0}{a_0} = \frac{dt_1}{a_1}. \tag{9.58}$$

The ratio of the wavelengths is given by

$$\frac{\lambda_0}{\lambda_1} = \frac{dt_0}{dt_1} = \frac{a_0}{a_1} \tag{9.59}$$

and the *cosmological redshift* z is defined via the ratio of the frequencies, that is,

$$1 + z = \frac{\nu_1}{\nu_0} = \frac{\lambda_0}{\lambda_1}. \tag{9.60}$$

Thus, we have

$$1 + z = \frac{a(t_0)}{a(t_1)}, \tag{9.61}$$

where t_0 (respectively, t_1) is the time the (first) light beam is received (respectively, emitted) by the corresponding observer.

Assuming finally that these times are close together, we have

$$1 + z = \frac{a(t_1 + dt)}{a(t_1)} \approx \frac{a(t_1) + \dot{a}(t_1)\, dt}{a(t_1)} = 1 + \left.\frac{\dot{a}}{a}\right|_{t_1} dt. \tag{9.62}$$

This is *Hubble's Law*, which says that the redshift z of a galaxy a distance d away is given approximately by

$$z \approx \frac{\dot{a}}{a}\, d, \tag{9.63}$$

since the time it takes to receive the light beam is the same as the distance to the object. The (present value of the) "constant" $\dot{a}/a$ is called *Hubble's constant.*

9.11 PROBLEMS

1. **Einstein Scalar**

   Using the relationship

   $$G^i{}_j = R^i{}_j - \frac{1}{2}\,\delta^i{}_j\, R,$$

   find an expression for the "Einstein scalar" $G = G^i{}_i$ in terms of the Ricci scalar $R = R^i{}_i$.

   (a) Determine the Ricci scalar for the Robertson–Walker line element.

   (b) Can a vacuum solution of Einstein's equation (with zero cosmological constant) have $R \neq 0$?

2. **Static Models**

The general Robertson–Walker (isotropic, homogeneous) cosmological model has line element

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2 \right)$$

with $k = -1, 0, 1$. The only nonzero components of the Einstein tensor are

$$G^t{}_t = -3\,\frac{\dot{a}^2 + k}{a^2}, \qquad G^r{}_r = G^\theta{}_\theta = G^\phi{}_\phi = -\,\frac{2a\ddot{a} + \dot{a}^2 + k}{a^2}.$$

(a) Assuming that the energy-momentum tensor is that of a perfect fluid

$$T^t{}_t = -\rho, \qquad T^r{}_r = T^\theta{}_\theta = T^\phi{}_\phi = p$$

with energy density $\rho > 0$ and pressure density $p \geq 0$, and using Einstein's field equation with cosmological constant Λ, namely

$$G^i{}_j + \Lambda\, \delta^i{}_j = 8\pi\, T^i{}_j,$$

show that a static model ($a =$ constant) is only possible if $k = 1$.

(b) Find the corresponding values of a and Λ in terms of ρ and p.

(c) How is your answer affected if ρ is allowed to be 0?

3. **The Standard Models**

(a) Consider the dust-filled ($p = 0$), closed ($k = 1$) Robertson–Walker cosmology, for which the Einstein equation (*without* cosmological constant) takes the form

$$\frac{3(\dot{a}^2 + 1)}{a^2} = 8\pi\rho; \qquad \rho a^3 = \frac{3c}{8\pi} = \text{constant}.$$

A parametric solution of these equations is given by

$$a = \frac{1}{2}c(1 - \cos\eta); \qquad t = \frac{1}{2}c(\eta - \sin\eta).$$

Draw a rough graph of a vs. t for $\eta \in [0, 2\pi]$.

(b) Consider the radiation-filled ($p = \rho/3$), closed ($k = 1$) Robertson–Walker cosmology, for which the Einstein equation (*without* cosmological constant) takes the form

$$\frac{3(\dot{a}^2 + 1)}{a^2} = 8\pi\rho; \qquad \rho a^4 = \frac{3b^2}{8\pi} = \text{constant}.$$

A parametric solution of these equations is given by

$$a = b\sin\eta; \qquad t = b(1-\cos\eta).$$

Draw a rough graph of a vs. t for $\eta \in [0, \pi]$.

(c) In each case, express the line element in terms of the variables $(\eta, \psi, \theta, \phi)$, where $r = \sin\psi$.

(d) By considering a radial (θ = constant, ϕ = constant), (initially) outgoing (r increasing), future pointing (t increasing), null curve in these two cosmologies, or otherwise, show that a beam of light emitted radially from $\psi = 0$ at $t = 0$ goes precisely once around the dust-filled universe (a) during its lifetime, and precisely halfway around the radiation-filled universe (b).

(e) Does this mean that one could see the back of one's head in either or both of these models?

Chapter 10

Solar System Applications

10.1 Bending of Light

Recall from Section 3.7 that the null geodesics of the Schwarzschild geometry satisfy

$$\dot{\phi} = \frac{\ell}{r^2}, \tag{10.1}$$

$$\dot{t} = \frac{e}{\left(1 - \frac{2m}{r}\right)}, \tag{10.2}$$

$$\dot{r}^2 = e^2 - \left(1 - \frac{2m}{r}\right)\frac{\ell^2}{r^2}. \tag{10.3}$$

Setting

$$u = \frac{1}{r}, \tag{10.4}$$

we have

$$\dot{r} = -\frac{1}{u^2}\dot{u} = -r^2\dot{u} = -\ell\,\frac{\dot{u}}{\dot{\phi}} = -\ell\,\frac{du}{d\phi} \tag{10.5}$$

so that the last equation in (10.3) becomes

$$\begin{aligned}\left(\frac{du}{d\phi}\right)^2 &= \frac{e^2}{\ell^2} - (1 - 2mu)\,u^2 \\ &= \frac{e^2}{\ell^2} - u^2 + 2mu^3.\end{aligned} \tag{10.6}$$

Differentiating (10.6) with respect to ϕ results in

$$\frac{d^2u}{d\phi^2} = -u + 3mu^2, \tag{10.7}$$

which we rewrite in the form

$$\frac{d^2(mu)}{d\phi^2} + mu = 3\,(mu)^2. \tag{10.8}$$

The right-hand side of (10.8) is a relativistic correction to the corresponding Newtonian equation.

Assuming $mu \ll 1$, we seek approximate solutions of (10.7) of the form

$$u = u_0 + u_1\epsilon, \tag{10.9}$$

where ϵ is of order mu. Substituting (10.9) into (10.8) and collecting terms of the same order yields

$$\frac{d^2(mu_0)}{d\phi^2} + mu_0 = 0, \tag{10.10}$$

$$\frac{d^2(mu_1)}{d\phi^2}\,\epsilon + mu_1\epsilon = 3(mu_0)^2. \tag{10.11}$$

The solution of the first equation is the straight line

$$u = \frac{1}{R}\sin\phi, \tag{10.12}$$

where R denotes the distance of closest approach to the origin, so that we could have taken

$$\epsilon = \frac{m}{R}. \tag{10.13}$$

A particular solution of the second equation is then

$$u_1\epsilon = \frac{m}{R^2}(1 + \cos^2\phi) \tag{10.14}$$

so that our approximate solution for u takes the form

$$u = \frac{1}{R}\sin\phi + \frac{m}{R^2}(1 + \cos^2\phi). \tag{10.15}$$

So our null geodesic is almost straight (the first term), and should in any case be nearly straight for r large, that is for u small. Noting that (10.15) implies that $\sin\phi < 0$, the (small, positive) asymptotic values $\phi_1 = 2\pi - \phi$ (on the right) and $\phi_2 = \phi - \pi$ (on the left) as r approaches ∞ give the (asymptotic) orientation of the line, as shown in Figure 10.1. Since u approaches zero as r approaches ∞, we must have

$$-\frac{\phi_i}{R} + 2\frac{m}{R^2} = 0 \tag{10.16}$$

and the deflection angle is just the sum

$$\delta = \phi_1 + \phi_2 = 4\frac{m}{R} = 4\frac{GM}{c^2R}. \tag{10.17}$$

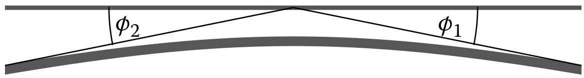


FIGURE 10.1. The bending of null geodesics. The horizontal line represents the Newtonian solution (10.12), which is a straight line, and the heavy line shows the approximate relativistic solution (10.15).

Since the mass of the sun is

$$m = 1.9891 \times 10^{33} \text{ g} \tag{10.18}$$

and its radius is

$$R = 6.955 \times 10^5 \text{ km}, \tag{10.19}$$

a light ray that just grazes the surface of the sun would be deflected by 1.75 seconds of arc, as shown schematically in Figure 10.2. Since this deflection is difficult to measure when the sun is shining, the best time to test this prediction of general relativity is during a total eclipse of the sun. The first expedition to attempt this was led by Sir Arthur Eddington in 1919; his observations confirmed Einstein's prediction and helped make Einstein famous.

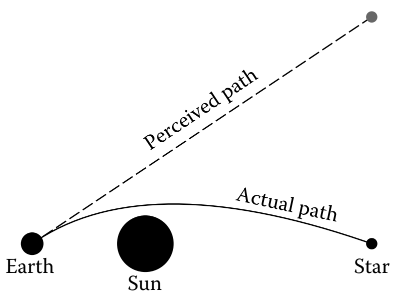


FIGURE 10.2. The bending of light by the sun (not to scale).

10.2 PERIHELION SHIFT OF MERCURY

Recall from Section 3.3 that the timelike geodesics of the Schwarzschild geometry satisfy

$$\dot{\phi} = \frac{\ell}{r^2}, \tag{10.20}$$

$$\dot{t} = \frac{e}{\left(1 - \frac{2m}{r}\right)}, \tag{10.21}$$

$$\dot{r}^2 = e^2 - \left(1 - \frac{2m}{r}\right)\left(1 + \frac{\ell^2}{r^2}\right). \tag{10.22}$$

Setting

$$u = \frac{1}{r}, \tag{10.23}$$

we have

$$\dot{r} = -\frac{1}{u^2}\dot{u} = -r^2\dot{u} = -\ell\,\frac{\dot{u}}{\dot{\phi}} = -\ell\,\frac{du}{d\phi} \tag{10.24}$$

so that the last equation in (10.22) becomes

$$\begin{aligned}\left(\frac{du}{d\phi}\right)^2 &= \frac{e^2}{\ell^2} - (1 - 2mu)\left(\frac{1}{\ell^2} + u^2\right)\\ &= \frac{e^2 - 1}{\ell^2} + \frac{2mu}{\ell^2} - u^2 + 2mu^3.\end{aligned} \tag{10.25}$$

Differentiating (10.25) with respect to ϕ results in

$$\frac{d^2u}{d\phi^2} = \frac{m}{\ell^2} - u + 3mu^2, \tag{10.26}$$

which we rewrite in the form

$$\frac{d^2(\ell u)}{d\phi^2} + \ell u = \frac{m}{\ell} + 3\,\frac{m}{\ell}\,(\ell u)^2. \tag{10.27}$$

The last term in (10.27) is a relativistic correction to the corresponding Newtonian equation and is small compared to the other terms for planetary motion.

Assuming $m^2 \ll \ell^2$, we seek approximate solutions of (10.26) of the form

$$u = \frac{m}{\ell}\left(u_0 + u_1\frac{m^2}{\ell^2}\right), \tag{10.28}$$

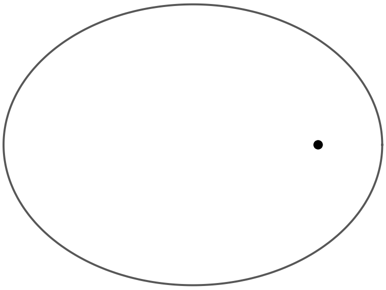


Figure 10.3. An ideal Newtonian orbit is an ellipse.

where we ignore higher powers of $\frac{m^2}{\ell^2}$. Substituting (10.28) into (10.27) and collecting terms of the same order yields

$$\frac{d^2(\ell u_0)}{d\phi^2} + \ell u_0 = 1, \tag{10.29}$$

$$\frac{d^2(\ell u_1)}{d\phi^2} + \ell u_1 = 3(\ell u_0)^2. \tag{10.30}$$

The solution of the first equation is the conic section

$$\ell u_0 = 1 + e\cos\phi \tag{10.31}$$

with eccentricity e. Idealized Newtonian orbits are indeed ellipses, as shown in Figure 10.3; for planetary orbits, the sun is at one of the foci of the ellipse, as shown. A particular solution of the second equation is given by

$$\ell u_1 = 3\left(1 + \frac{e^2}{2} + e\phi\sin\phi - \frac{e^2}{6}\cos(2\phi)\right). \tag{10.32}$$

The physically interesting term in (10.32) is the third, which continues to increase with each revolution. Ignoring the other corrections, we have

$$\begin{aligned} u &\approx \frac{m}{\ell^2}\left(1 + e\cos\phi + \frac{3em^2}{\ell^2}\phi\sin\phi\right) \\ &\approx \frac{m}{\ell^2}\left(1 + e\cos\left(\phi\left(1 - \frac{3m^2}{\ell^2}\right)\right)\right). \end{aligned} \tag{10.33}$$

Thus, the period is no longer 2π, but rather

$$\frac{2\pi}{1 - 3\frac{m^2}{\ell^2}} \approx 2\pi\left(1 + 3\frac{m^2}{\ell^2}\right). \tag{10.34}$$

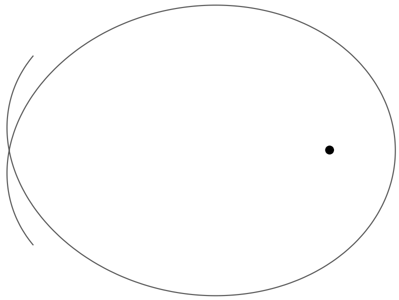


FIGURE 10.4. A schematic representation of the relativistic perihelion shift.

The *perihelion shift*, shown schematically in Figure 10.4, is therefore given by

$$\delta\phi = 2\pi - 2\pi\left(1 + 3\frac{m^2}{\ell^2}\right) = 6\pi\frac{m^2}{\ell^2} = 6\pi\left(\frac{Gm}{c\ell}\right)^2 \tag{10.35}$$

in radians per orbit, where we have restored the Newtonian conversion factors in the last step.

To this accuracy, we can determine the angular momentum ℓ using the Newtonian orbit. The mass of the sun is

$$m = 1.9891 \times 10^{33}\ \mathrm{g} \tag{10.36}$$

and the angular momentum can be computed from the Newtonian formula

$$\ell^2 = GMa(1 - e^2), \tag{10.37}$$

where a is the length of the semimajor axis of the elliptical orbit. Inserting astronomical data for the orbit of Mercury, namely

$$a = 57.9 \times 10^6\ \mathrm{km}, \tag{10.38}$$

$$e = 0.2056, \tag{10.39}$$

into (10.35), and using the fact that the orbital period of Mercury is

$$T = 87.97\ \text{days}, \tag{10.40}$$

one obtains a perihelion shift of

$$\frac{100\ \text{years}}{T}\,\delta\phi = 42.98\ \frac{\text{seconds of arc}}{\text{century}}. \tag{10.41}$$

Thus, the rate of precession of the perihelion of the orbit of Mercury is just under 43 seconds of arc per century—precisely the discrepancy between observation and the Newtonian computation.

10.3 Global Positioning

Global positioning systems compare signals from multiple satellites in order to determine the position of an object. For the computation to be accurate, one must be able to compare the rate at which signals are emitted with the rate at which they are received. Equivalently, the clocks on the satellites must be synchronized with clocks on the ground. This synchronization turns out to require both special and general relativity!

We present here an order-of-magnitude computation to illustrate the need for relativistic corrections. Let dt denote an infinitesimal time interval on the Earth's surface, and let $d\tau$ denote an infinitesimal time interval on the satellite. These two quantities will differ due to the time dilation between moving reference frames (special relativity), as well as due to the difference in the gravitational field at different altitudes (general relativity).

We first consider time dilation. The speed of a typical satellite is given by

$$\tanh\beta = \frac{v}{c} \approx 1.3 \times 10^{-5}. \tag{10.42}$$

Time dilation leads to an Earth observer thinking that the satellite clock runs slow by a factor of $\cosh\beta$, that is,

$$d\tau = \frac{1}{\cosh\beta}\, dt. \tag{10.43}$$

But for such small speeds we have

$$\frac{1}{\cosh\beta} = \sqrt{1-\tanh^2\beta} \approx 1 - \frac{1}{2}\tanh^2\beta \approx 1 - 0.8 \times 10^{-10}. \tag{10.44}$$

In other words, special relativity leads to a time dilation effect of roughly one part in 10^{10}.

Earth's gravitational field can be modeled by the Schwarzschild metric; dt now refers to "far-away" time, which is a good approximation to time on the ground. Ignoring the satellite's motion, the relevant part of the line element is just

$$d\tau \approx \sqrt{1-\frac{2m}{r}}\, dt. \tag{10.45}$$

Estimating Earth's mass as 4 mm, its radius as 6000 km, and the radius of the satellite orbit as 26,000 km, we obtain

$$\sqrt{1-\frac{2m}{r}} \approx 1 - \frac{m}{r} \approx 1 - 1.7 \times 10^{-10}, \tag{10.46}$$

which is larger than the special relativistic effect, but of the same order of magnitude.

Why do errors on the order of one part in 10^{10} matter? Because they accumulate. There are roughly 10^5 seconds in a day, so the daily error is roughly 10^4 nanoseconds. Although a few nanoseconds might not seem like much, light travels on the order of 1 km in that amount of time. If relativistic effects, both special and general, were not taken into account, the resulting error in location would increase by roughly 1 km per day, which is definitely enough to matter.

The approximations used here are rather rough, not only in terms of the accuracy of the data used, but also in the assumptions made. There is a difference between clocks at infinity and clocks on the Earth's surface, the Earth's rotation must be taken into account, there are Doppler effects, and so forth. Nonetheless, the principle is sound: general relativity matters to global positioning systems.

Part III

Differential Forms

≻ Chapter 11 ≺

Calculus Revisited

11.1 Differentials

Differentiation is about small changes. We interpret the *differential* df as the "small change in f," so that the basic differentiation operation becomes

$$f \longmapsto df, \tag{11.1}$$

which we describe as "zapping f with d."

The derivative rules from first-term calculus can be written in differential form as

$$d(u^n) = nu^{n-1}\,du, \tag{11.2}$$
$$d(e^u) = e^u\,du, \tag{11.3}$$
$$d(\sin u) = \cos u\,du, \tag{11.4}$$
$$d(uv) = u\,dv + v\,du, \tag{11.5}$$
$$d(u + cv) = du + c\,dv, \tag{11.6}$$

where c is a constant. The first three rules suffice to differentiate all elementary functions,[1] and the latter two, referred to as the *Leibniz property* and *linearity*, respectively, are distinguishing characteristics of many different notions of differentiation. The quotient rule is not needed, as it follows from the product rule together with the power rule. Derivatives of inverse functions are most easily computed directly, for example,

$$u = \ln v \Longrightarrow v = e^u \Longrightarrow dv = e^u\,du = v\,du \Longrightarrow du = \frac{dv}{v}. \tag{11.7}$$

Finally, the chain rule is conspicuous by its absence. For example, if $f = q^2$ and $q = \sin u$, then

$$df = 2q\,dq = 2\sin u \cos u\,du \tag{11.8}$$

[1] A separate rule is not needed for $\cos u = \sin(\frac{\pi}{2} - u)$. The rule for $\sin u$ could also be eliminated, using the Euler formula $e^{iu} = \cos u + i \sin u$.

so that

$$\frac{df}{du} = \frac{df}{dq}\frac{dq}{du} \tag{11.9}$$

so that the traditional chain rule is built into differential notation. Put differently, we have

$$df = \frac{df}{dq}\,dq = \frac{df}{du}\,du. \tag{11.10}$$

The differential df is not itself a derivative, as we have not yet specified "with respect to what." Rather, the derivative of f with respect to, say, u is just the ratio of the small changes df and du.

Functions of several variables can be differentiated without further ado. For example, suppose that $f = \sin u$ with $u = pq$. Using the product rule, we obtain

$$df = \cos u\,du = \cos u(p\,dq + q\,dp) = \cos(pq)(p\,dq + q\,dp). \tag{11.11}$$

It is important to realize that differentials are not themselves the answers to any physical questions. So what was the question? Perhaps the goal was to compute the derivative of f with respect to p. Strictly speaking, this question is poorly posed, and one should seek the derivative of f with respect to *p with q held constant.* But that's easy: If q is constant, $dq = 0$, and the desired derivative is just the coefficient of dp, written

$$\frac{\partial f}{\partial p} = q\,\cos(pq). \tag{11.12}$$

We have expressed df in terms of dp and dq; this is very much like finding the components of a vector ("df") in terms of a basis ("$\{dp, dq\}$"). Furthermore, the "components" are just partial derivatives of f:

$$df = \frac{\partial f}{\partial p}\,dp + \frac{\partial f}{\partial q}\,dq. \tag{11.13}$$

This interplay between calculus and algebra is at the heart of the study of differential forms.

11.2 INTEGRANDS

Integration is about chopping a region into small pieces and then adding up some small quantity on each piece. Thus, integration is about adding up differentials, and the basic integration operation is

$$f = \int df. \tag{11.14}$$

In single-variable calculus, such integrals take the form

$$W = \int F\,dx, \tag{11.15}$$

which might represent the work done by the force F when moving an object in the x-direction. The *integrand* in this case is $F\,dx$, where F is a function of x. Similarly, a typical *double integral* takes the form $\int F\,dx\,dy$, where now F depends on x and y; the integrand in this case is $F\,dx\,dy$.

Thus, differentials, and products of differentials, are the things one integrates!

Differential forms are just integrands, together with rules for manipulating them, using both algebra and calculus.

11.3 CHANGE OF VARIABLES

Consider a change of variables in two dimensions

$$x = x(u,v), \qquad y = y(u,v). \tag{11.16}$$

This is just a special case of a parametric surface!

To find the surface element of such surfaces, first foliate the surface with curves along which either u or v is constant. We have

$$d\vec{\boldsymbol{r}} = \frac{\partial\vec{\boldsymbol{r}}}{\partial u}\,du + \frac{\partial\vec{\boldsymbol{r}}}{\partial v}\,dv \tag{11.17}$$

and only one of these terms is nonzero along each such curve. The surface element can now be obtained as

$$d\vec{\boldsymbol{A}} = d\vec{\boldsymbol{r}}_1 \times d\vec{\boldsymbol{r}}_2 = \left(\frac{\partial\vec{\boldsymbol{r}}}{\partial u} \times \frac{\partial\vec{\boldsymbol{r}}}{\partial v}\right) du\,dv. \tag{11.18}$$

We have

$$\frac{\partial\vec{\boldsymbol{r}}}{\partial u} \times \frac{\partial\vec{\boldsymbol{r}}}{\partial v} = \begin{vmatrix} \hat{\boldsymbol{x}} & \hat{\boldsymbol{y}} & \hat{\boldsymbol{z}} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & 0 \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & 0 \end{vmatrix} = \hat{\boldsymbol{z}} \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix}. \tag{11.19}$$

This determinant is important enough to have its own name; we define the *Jacobian* of the transformation from u, v to x, y to be[2]

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}. \tag{11.20}$$

[2]Recall that the determinant of the transpose of a matrix is the same as the determinant of the matrix.

Since $d\vec{\boldsymbol{A}} = dx\,dy\,\hat{\boldsymbol{z}}$ for *any* (part of) the plane (oriented upward), we obtain

$$dx\,dy = \frac{\partial(x,y)}{\partial(u,v)}\,du\,dv. \tag{11.21}$$

Unlike standard usage in multivariable calculus, we do not take the absolute value of the Jacobian determinant; we care about the relative orientation of the two sets of coordinates.

Not surprisingly, when this procedure is applied to the polar coordinate transformation

$$x = r\cos\phi, \qquad y = r\sin\phi, \tag{11.22}$$

we get (check this yourself!)

$$\frac{\partial(x,y)}{\partial(r,\phi)} = r \tag{11.23}$$

so that, as we already knew, $dx\,dy = r\,dr\,d\phi$.

In practice, one is often given u and v in terms of x and y, not the other way around. Rather than solving for x and y, which can be difficult, there is a simpler way. Inverting the roles of the two sets of variables in (11.20), we must have

$$du\,dv = \frac{\partial(u,v)}{\partial(x,y)}\,dx\,dy \tag{11.24}$$

from which we can conclude that

$$\frac{\partial(x,y)}{\partial(u,v)} = 1 \bigg/ \frac{\partial(u,v)}{\partial(x,y)}. \tag{11.25}$$

11.4 MULTIPLYING DIFFERENTIALS

We are now ready to multiply differentials. Starting from

$$du\,dv = \frac{\partial(u,v)}{\partial(x,y)}\,dx\,dy, \tag{11.26}$$

first set $u = y$ and $v = x$ to obtain

$$dy\,dx = -dx\,dy. \tag{11.27}$$

Thus, multiplication of differentials is antisymmetric. Now set $u = v = x$,

in which case the determinant is 0, so that

$$dx\,dx = 0. \tag{11.28}$$

Thus, differentials "square" to zero.

In fact, these two properties are equivalent. Clearly, antisymmetry implies that squares are zero, since only zero is unchanged when multiplied by -1. But the reverse is also true: If all squares are zero, then from

$$(dx + dy)(dx + dy) = 0 \tag{11.29}$$

and $dx\,dx = 0 = dy\,dy$ we obtain

$$dx\,dy + dy\,dx = 0 \tag{11.30}$$

from which antisymmetry follows. (This technique is called *polarization*.)

≻ Chapter 12 ≺

Vector Calculus Revisited

12.1 A Review of Vector Calculus

The basic objects in vector calculus are vector fields $\vec{\boldsymbol{F}}$ in three-dimensional Euclidean space, $\mathbb{R}^3$. A vector field can be expressed in terms of a *basis* by giving its *components* with respect to that basis, such as[1]

$$\vec{\boldsymbol{F}} = F_x\,\hat{\boldsymbol{x}} + F_y\,\hat{\boldsymbol{y}} + F_z\,\hat{\boldsymbol{z}}, \tag{12.1}$$

where $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ denotes the standard rectangular basis for $\mathbb{R}^3$, also written as $\{\hat{\boldsymbol{\imath}}, \hat{\boldsymbol{\jmath}}, \hat{\boldsymbol{k}}\}$. See Chapter 16 for a more detailed treatment.

Integration

One of the two fundamental operations in calculus is integration. In vector calculus, we are interested in integrating vector fields and can do so over a *curve* C,

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}},$$

which might represent the work done by the force $\vec{\boldsymbol{F}}$ when moving an object along the path C, or over a surface S,

$$\int_S \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}},$$

which might represent the flux of $\vec{\boldsymbol{F}}$ through S. Expressions for $d\vec{\boldsymbol{r}}$ in various coordinate systems can be computed explicitly by considering infinitesimal vector displacements; in rectangular coordinates, we have

$$d\vec{\boldsymbol{r}} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}} + dz\,\hat{\boldsymbol{z}}. \tag{12.2}$$

Similarly, coordinate expressions for $d\vec{\boldsymbol{A}}$ can be computed by chopping up the surface into infinitesimal parallelograms, whose (directed) areas are just

[1] We use subscripts to denote components, *not* partial differentiation.

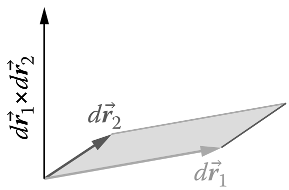


FIGURE 12.1. Infinitesimal area expressed as a cross product.

the cross product of their sides, so that $d\vec{\boldsymbol{A}}$ can be expressed in the form

$$d\vec{\boldsymbol{A}} = d\vec{\boldsymbol{r}}_1 \times d\vec{\boldsymbol{r}}_2 \tag{12.3}$$

as shown in Figure 12.1.

Vector calculus also involves the integration of functions over a region R,

$$\int_R f\,dV,$$

which might represent the total amount of chocolate in R if f gives the density of chocolate. The volume element can be determined by chopping up the region into infinitesimal parallelepipeds, whose volumes can be found using the triple product, leading to an expression of the form

$$dV = d\vec{\boldsymbol{r}}_1 \times d\vec{\boldsymbol{r}}_2 \cdot d\vec{\boldsymbol{r}}_3 \tag{12.4}$$

as shown in Figure 12.2. Finally, we can evaluate functions at a point P, and we extend the ordinary notions of integration to include this case by defining

$$\int_P f = f\Big|_P. \tag{12.5}$$

Thus, we can integrate over zero-, one-, two-, or three-dimensional regions.

DIFFERENTIATION

The other fundamental operation in calculus is differentiation. Much of vector calculus involves vector derivative operators, namely the *gradient* of a function ($\vec{\boldsymbol{\nabla}} f$), the *curl* of a vector field ($\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}$), and the *divergence* of a

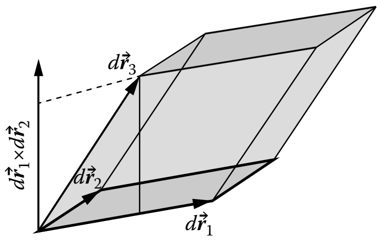


FIGURE 12.2. Infinitesimal volume expressed as a triple product.

vector field ($\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}$). In rectangular coordinates, these operations are given by

$$\vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial x}\,\hat{\boldsymbol{x}} + \frac{\partial f}{\partial y}\,\hat{\boldsymbol{y}} + \frac{\partial f}{\partial z}\,\hat{\boldsymbol{z}}, \tag{12.6}$$

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} = \begin{vmatrix} \hat{\boldsymbol{x}} & \hat{\boldsymbol{y}} & \hat{\boldsymbol{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix}, \tag{12.7}$$

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \tag{12.8}$$

and satisfy the identities

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{\nabla}} f = \vec{\boldsymbol{0}}, \tag{12.9}$$

$$\vec{\boldsymbol{\nabla}} \cdot \left(\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}\right) = 0. \tag{12.10}$$

THEOREMS

Each of the above derivative operations leads to a version of the fundamental theorem of calculus. Integrating a gradient along a curve C from point A to point B results in

$$\int_C \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} = f\Big|_A^B, \tag{12.11}$$

which could be called the *fundamental theorem for the gradient*. Integrating a curl over a surface S with boundary C leads to

$$\int_S \left(\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}\right) \cdot d\vec{\boldsymbol{A}} = \oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}, \tag{12.12}$$

(with appropriate orientations) which is *Stokes' Theorem*. Finally, integrating a divergence over a region R with boundary S leads to

$$\int_R \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}} \, dV = \oint_S \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}, \tag{12.13}$$

which is the *Divergence Theorem*, and which relates the divergence of $\vec{\boldsymbol{F}}$ *inside* S to the flux of $\vec{\boldsymbol{F}}$ *out of* S.

12.2 DIFFERENTIAL FORMS IN THREE DIMENSIONS

Differential forms in three (Euclidean) dimensions are just the integrands discussed in Section 12.1.[2] A *1-form* F is the integrand of a *line integral*, that is,

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}, \tag{12.14}$$

where we deliberately use the same base letter F for the vector field and the corresponding 1-form. A *2-form* α is the integrand of a *surface integral*, that is,

$$\alpha = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}, \tag{12.15}$$

and a *3-form* β is the integrand of a *volume integral*, that is,

$$\beta = f \, dV. \tag{12.16}$$

Finally, a *0-form* f is the "integrand" of the zero-dimensional integral defined in Section 12.1, namely f itself.

Examples of 0-forms are easy: Let f be any (reasonable) function. Each 0-form (function) f also defines a 3-form, which in rectangular coordinates takes the form

$$\beta = f \, dx \, dy \, dz. \tag{12.17}$$

Each vector field $\vec{\boldsymbol{F}}$ defines a 1-form, which in rectangular coordinates takes the form

$$F = F_x \, dx + F_y \, dy + F_z \, dz. \tag{12.18}$$

But what are 2-forms?

Suppose the surface S is the xy-plane, given by $z = 0$. Then clearly

$$d\vec{\boldsymbol{A}} = dx \, \hat{\boldsymbol{x}} \times dy \, \hat{\boldsymbol{y}} = dx \, dy \, \hat{\boldsymbol{z}} \tag{12.19}$$

[2]There are other integrals in vector calculus, such as the total amount of chocolate on a surface or curve, but such integrals cannot be written in terms of differential forms, at least not in three dimensions.

so that

$$\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = F_z\, dx\, dy. \tag{12.20}$$

Repeating this argument for the other two coordinate planes correctly suggests that the 2-form defined by $\vec{\boldsymbol{F}}$ is given by

$$\alpha = F_x\, dy\, dz + F_y\, dz\, dx + F_z\, dx\, dy. \tag{12.21}$$

12.3 Multiplication of Differential Forms

We have seen that the *rank* p of differential forms in $\mathbb{R}^3$ can be 0, 1, 2, or 3; the rank "counts" the number of times that "d" appears in each term. So 2-forms and 3-forms involve products of 1-forms; differential forms form an *algebra* and can be multiplied. What are the rules?

We have been somewhat casual so far in our treatment of vector integration. In particular, we haven't yet worried about the *orientation* of the surfaces over which we integrate. It is clear, however, that the value of a vector line integral ("work") changes sign if we traverse the curve backward. Similarly, a vector surface integral ("flux") changes sign if we choose the opposite normal vector, that is, if we use the opposite orientation of $d\vec{\boldsymbol{A}}$. So it's not quite correct to say that

$$\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = F_z\, dx\, dy \tag{12.22}$$

as we did in Section 12.2. The vector surface element $d\vec{\boldsymbol{A}}$ carries with it a choice of orientation. For instance, in the xy-plane, we must distinguish between $dx\,\hat{\boldsymbol{x}} \times dy\,\hat{\boldsymbol{y}}$ and $dy\,\hat{\boldsymbol{y}} \times dx\,\hat{\boldsymbol{x}}$, which differ by a sign, whereas $dx\,dy$ and $dy\,dx$ are normally regarded as being equal.[3] We therefore *define* multiplication of differential forms to be antisymmetric, and from now on use the symbol $\wedge$, read as "wedge," to distinguish this product from ordinary multiplication. We have

$$dy \wedge dx = -dx \wedge dy \tag{12.23}$$

by definition, and we conventionally define

$$\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = F_x\, dy \wedge dz + F_y\, dz \wedge dx + F_z\, dx \wedge dy. \tag{12.24}$$

Similarly, we regard volume integrals as being dependent on a choice of orientation ("right-handed" or "left-handed"), and regard dV as representing the (standard, right-handed) orientation given by

$$dV = dx \wedge dy \wedge dz. \tag{12.25}$$

[3]This argument can be made more precise using Jacobians.

12.4 RELATIONSHIPS BETWEEN DIFFERENTIAL FORMS

There is clearly a relationship between 0-forms and 3-forms, given by relating f and $f\,dV$. We can therefore define a map from 0-forms to 3-forms, which we write as

$$*f = f\,dV. \tag{12.26}$$

Similarly, relating $\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$ and $\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}$ yields a relationship between 1-forms and 2-forms, which we write as

$$*(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}) = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}. \tag{12.27}$$

In three (Euclidean) dimensions, we also write $*$ for the inverse maps, so that

$$*(f\,dV) = f \tag{12.28}$$

and

$$*(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}) = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}. \tag{12.29}$$

We can work out the action of the *Hodge dual* map $*$ on a basis at each rank. The standard basis of 1-forms in $\mathbb{R}^3$ is the set $\{dx, dy, dz\}$; any 1-form can be expressed as a linear combination of these basis 1-forms, with coefficients that are *functions*. Similarly, the standard basis of 2-forms is $\{dy \wedge dz, dz \wedge dx, dx \wedge dy\}$, and the standard basis of 3-forms is $\{dx \wedge dy \wedge dz\}$. Finally, the standard basis of 0-forms is just $\{1\}$, as any function is a multiple of the constant function 1 (with a coefficient that is a function). The action of $*$ is therefore given by

$$*1 = dx \wedge dy \wedge dz, \tag{12.30}$$
$$*dx = dy \wedge dz, \tag{12.31}$$
$$*dy = dz \wedge dx, \tag{12.32}$$
$$*dz = dx \wedge dy, \tag{12.33}$$
$$*(dy \wedge dz) = dx, \tag{12.34}$$
$$*(dz \wedge dx) = dy, \tag{12.35}$$
$$*(dx \wedge dy) = dz, \tag{12.36}$$
$$*(dx \wedge dy \wedge dz) = 1, \tag{12.37}$$

along with the linearity property

$$*(f\,\alpha + \beta) = f *\alpha + *\beta \tag{12.38}$$

for any function f and any p-forms α, β.

12.5 DIFFERENTIATION OF DIFFERENTIAL FORMS

How do we differentiate differential forms?

It's pretty clear how to differentiate 0-forms (functions): Take the gradient. The 1-form associated with $\vec{\boldsymbol{\nabla}} f$ is

$$df = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}}, \tag{12.39}$$

which we sometimes refer to as the *Master Formula* in vector calculus.

Our remaining vector derivative operators are the curl, which takes vector fields to vector fields, and the divergence, which takes vector fields to scalar fields. We would like to express these operators in terms of differential forms. Since adding a "d" takes a p-form to a $(p+1)$-form, a reasonable choice is to *define* differentiation of 1- and 2-forms via

$$d\left(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}\right) = \left(\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}\right) \cdot d\vec{\boldsymbol{A}}, \tag{12.40}$$

$$d\left(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}\right) = \left(\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}\right) dV. \tag{12.41}$$

Using the relationship between vector fields $\vec{\boldsymbol{G}}$ and 1-forms $\vec{\boldsymbol{G}} \cdot d\vec{\boldsymbol{r}}$, it is now easy to check that $*dF$ is the 1-form associated with $\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}$, and $*d{*}F$ is the 0-form $\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}$. Explicitly, we have

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}, \tag{12.42}$$

$$dF = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}, \tag{12.43}$$

$$*dF = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} \tag{12.44}$$

and

$$*F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}, \tag{12.45}$$

$$d{*}F = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}\, dV, \tag{12.46}$$

$$*d{*}F = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}. \tag{12.47}$$

We can regard these relations as the definitions of the curl and divergence of a 1-form, namely

$$*dF = \operatorname{curl}(F), \tag{12.48}$$

$$*d{*}F = \operatorname{div}(F), \tag{12.49}$$

and of course we have

$$df = \text{grad}(f) \tag{12.50}$$

for 0-forms.[4]

Why have we rewritten vector calculus in this new language? Because it unifies the results. By convention, d acting on a 3-form is zero; the result should be a 4-form, but there are no (nonzero) 4-forms in three dimensions. Both second-derivative identities (12.9) and (12.10) therefore reduce to the single statement that

$$d^2 = 0 \tag{12.51}$$

when acting on p-forms for any p. Furthermore, all of the integral theorems of vector calculus can be combined into the single statement

$$\int_D d\alpha = \int_{\partial D} \alpha, \tag{12.52}$$

where D is a p-dimensional region in $\mathbb{R}^3$, α is a $(p-1)$-form, and ∂D denotes the boundary of D.

The remainder of this book seeks to generalize this language to other dimensions (and signatures).

[4]In casual usage, one often refers to 1-forms as "vectors," leading one to interpret F and df (and curl(F)) as "vectors" when they are really 1-forms. More precisely, if $F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$, then we have

$$\begin{aligned}
\text{curl}(F) &= \left(\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}\right) \cdot d\vec{\boldsymbol{r}}, \\
\text{div}(F) &= \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}, \\
\text{grad}(f) &= \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}}.
\end{aligned}$$

≻ Chapter 13 ≺

The Algebra of Differential Forms

13.1 Differential Forms

We are now ready to provide a definition of *differential forms*. Consider $\mathbb{R}^n$ with coordinates $\{x^i\}$, where $i = 1, ..., n$. (Any surface M in $\mathbb{R}^n$ can be used instead.) Now construct all linear combinations of the differentials $\{dx^i\}$, that is, consider the space V defined by

$$V = \left\langle \{dx^i\} \right\rangle = \{a_i\, dx^i\} \tag{13.1}$$

where we have introduced the *Einstein summation convention*, under which repeated indices must be summed over. If the coefficients a_i are *numbers*, then V is an n-dimensional vector space with basis $\{dx^i\}$. We will, however, instead allow the coefficients to be *functions* on $\mathbb{R}^n$, which turns V into a module over the ring of functions. (This is entirely analogous to the transition from *vectors* to *vector fields*.)

What are the elements of V? Any differential df can be expanded in terms of our basis using calculus, namely

$$df = \frac{\partial f}{\partial x^i}\, dx^i \tag{13.2}$$

so that $df \in V$. What integrand does df correspond to? The fundamental theorem for the gradient says that

$$\int_C \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} = f\Big|_A^B \tag{13.3}$$

for any curve C from point A to point B. We rewrite this relationship in terms of integrands as

$$df = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}}, \tag{13.4}$$

which we refer to as the *Master Formula* because of its importance in vector calculus. Thus, the differential form df represents the integrand corresponding to the vector field $\vec{\boldsymbol{\nabla}} f$; we will have more to say about this correspondence in Chapter 16.

A special case of (13.2) occurs when $f = x^j$ is a coordinate function, in which case

$$d(x^j) = \frac{\partial x^j}{\partial x^i}\, dx^i = dx^j, \tag{13.5}$$

which shows that our basis really is what we thought. And we can use calculus to check that this construction does not depend on the choice of coordinates, that is, that *df* is the same when expanded in terms of any set of coordinates on $\mathbb{R}^n$. Schematically, this argument goes as follows:

$$\begin{aligned} df &= \frac{\partial f}{\partial u}\, du + \frac{\partial f}{\partial v}\, dv + ... \\ &= \frac{\partial f}{\partial u}\left(\frac{\partial u}{\partial x}\, dx + ...\right) + \frac{\partial f}{\partial v}\left(\frac{\partial v}{\partial x}\, dx + ...\right) + ... \\ &= \left(\frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + ...\right) dx + ... \\ &= \frac{\partial f}{\partial x}\, dx + ... \;, \end{aligned} \tag{13.6}$$

which shows how to use calculus to convert between one basis and another.

We will refer to elements of V as *1-forms* on $\mathbb{R}^n$ and will henceforth write V itself as $\bigwedge^1(\mathbb{R}^n)$, which we will often abbreviate to $\bigwedge^1$.

We already know how to multiply 1-forms, which we now write with an explicit symbol, $\wedge$, read as "wedge": Use the rules for integrands. Thus, we require

$$\alpha \wedge \alpha = 0, \tag{13.7}$$
$$\beta \wedge \alpha = -\alpha \wedge \beta \tag{13.8}$$

for any $\alpha, \beta \in \bigwedge^1$. We also assume distributivity and associativity, so that

$$(\alpha + \beta) \wedge \gamma = \alpha \wedge \gamma + \beta \wedge \gamma, \tag{13.9}$$
$$f(\alpha \wedge \gamma) = (f\alpha) \wedge \gamma, \tag{13.10}$$
$$(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma) \tag{13.11}$$

for any $\alpha, \beta, \gamma \in \bigwedge^1$ and function f.

13.2 HIGHER-RANK FORMS

We refer to the product of two 1-forms as a *2-form*. The space of all 2-forms, denoted $\bigwedge^2(\mathbb{R}^n)$ or simply $\bigwedge^2$, is therefore spanned by all products of basis

1-forms, that is,

$$\bigwedge^2 = \left\langle \{dx^i \wedge dx^j\} \right\rangle . \tag{13.12}$$

But this set is redundant; it is sufficient to assume $i < j$. Similarly, we can define *p-forms* as

$$\bigwedge^p = \left\langle \{dx^{i_1} \wedge ... \wedge dx^{i_p}\} \right\rangle , \tag{13.13}$$

where $1 \le i_1 < ... < i_p \le n$.

Each of the spaces $\bigwedge^p$ is itself a vector space (or module), and since we have an explicit basis, we can determine its dimension, which is

$$\dim\left(\bigwedge^p\right) = \binom{n}{p} = \frac{n!}{p!(n-p)!} \tag{13.14}$$

at least so long as $p \le n$. What if $p = 0$? We define $\bigwedge^0$ to be the space of scalars (functions), with dimension 1. What if $p > n$? Since there are only n independent 1-forms, any product with more than n factors must be zero. Thus,

$$\dim\left(\bigwedge^p\right) = 0 \qquad\qquad (p > n). \tag{13.15}$$

13.3 POLAR COORDINATES

Consider $\mathbb{R}^2$ with the usual rectangular coordinates (x, y). The standard basis of 1-forms is $\{dx, dy\}$. But we could also use polar coordinates (r, ϕ), and the corresponding basis $\{dr, d\phi\}$. How are these bases related?

We have

$$x = r\,\cos\phi, \tag{13.16}$$

$$y = r\,\sin\phi, \tag{13.17}$$

which leads to

$$dx = dr\,\cos\phi - r\,\sin\phi\,d\phi, \tag{13.18}$$

$$dy = dr\,\sin\phi + r\,\cos\phi\,d\phi. \tag{13.19}$$

The inverse transformation can be obtained by inverting these expressions, or directly starting from

$$r^2 = x^2 + y^2, \tag{13.20}$$

$$\tan\phi = \frac{y}{x}, \tag{13.21}$$

which leads to

$$2r\,dr = 2x\,dx + 2y\,dy, \tag{13.22}$$

$$(1 + \tan^2 \phi)\,d\phi = \frac{x\,dy - y\,dx}{x^2}, \tag{13.23}$$

from which it follows that

$$r\,dr = x\,dx + y\,dy, \tag{13.24}$$

$$r^2\,d\phi = x\,dy - y\,dx. \tag{13.25}$$

What about 2-forms? There is only one independent 2-form, namely $dx \wedge dy$, and multiplying this out using (13.18) and (13.19) one obtains

$$dx \wedge dy = r\,dr \wedge d\phi. \tag{13.26}$$

Equivalently, multiplying (13.24) and (13.25) together yields

$$r^3\,dr \wedge d\phi = (x^2 + y^2)\,dx \wedge dy, \tag{13.27}$$

which is of course the same thing.

It is worth noting that in the language of differential forms these two "area elements" are identically equal, whereas the geometric areas "$dx\,dy$" and "$r\,dr\,d\phi$" in fact differ by second-order terms. The differential statement "$dx\,dy = r\,dr\,d\phi$" is really an assertion about integrals, not infinitesimal areas!

13.4 LINEAR MAPS AND DETERMINANTS

To demonstrate the power of differential forms, we consider some elementary applications. A map

$$A : V \longmapsto V \tag{13.28}$$

is *linear* if it satisfies[1]

$$A(f\alpha + \beta) = f(A\alpha) + A\beta. \tag{13.29}$$

A linear map on 1-forms can be extended to a linear map on p-forms, also called A (or occasionally $\bigwedge^p A$), given by

$$\alpha^1 \wedge \ldots \wedge \alpha^p \longmapsto A\alpha^1 \wedge \ldots \wedge A\alpha^p. \tag{13.30}$$

[1] In this section, it is convenient to work with vector spaces rather than modules, in which case f is a number rather than a function.

In particular, A induces a linear map on $\bigwedge^n(V)$, which is one-dimensional. But the only such map is multiplication by a scalar, which we call $|A|$. Thus, for any $\omega \in \bigwedge^n(V)$, we have

$$A(\omega) = |A|\,\omega. \tag{13.31}$$

The linear map A can be expressed in terms of a basis dx^i as

$$A(dx^i) = a^i{}_j dx^j, \tag{13.32}$$

where of course $\left(a^i{}_j\right)$ is precisely the matrix representation of A with respect to the given basis. Working out the action of A on the basis element $dx^1 \wedge ... \wedge dx^n$ of $\bigwedge^n(V)$, it is easily seen that $|A|$ consists of a sum of all possible products of one element from each row and column of the matrix $\left(a^i{}_j\right)$, with suitable signs; $|A|$ is the *determinant* of A. Note, however, that the definition (13.31) is independent of basis!

We can now give an elegant derivation of the product rule for the determinant. Suppose A and B are linear transformations on V, and ω is any element of $\bigwedge^n(V)$. Then the composition $A \circ B$ satisfies

$$\begin{aligned}|A \circ B|\,\omega = (A \circ B)\,\omega = A(B\,\omega) = A(|B|\,\omega)\\ = |B|(A\,\omega) = |B||A|\,\omega = |A||B|\,\omega\end{aligned} \tag{13.33}$$

so that the determinant of a product is the product of the determinants.

13.5 THE CROSS PRODUCT

Consider $\bigwedge^1(\mathbb{R}^3)$. Since all three-dimensional vector spaces are isomorphic, we can identify the 1-form

$$v = v_x\,dx + v_y\,dy + v_z\,dz \tag{13.34}$$

with the ordinary vector (field)

$$\vec{\boldsymbol{v}} = v_x\,\hat{\boldsymbol{x}} + v_y\,\hat{\boldsymbol{y}} + v_z\,\hat{\boldsymbol{z}} \tag{13.35}$$

and similarly for

$$w = w_x\,dx + w_y\,dy + w_z\,dz. \tag{13.36}$$

In other words, we identify the basis vectors as follows:

$$\hat{\boldsymbol{x}} \longleftrightarrow dx, \tag{13.37}$$

$$\hat{\boldsymbol{y}} \longleftrightarrow dy, \tag{13.38}$$

$$\hat{\boldsymbol{z}} \longleftrightarrow dz. \tag{13.39}$$

This identification turns vectors into things we can integrate and can be thought of as the map from vectors to 1-forms given by

$$v = \vec{\boldsymbol{v}} \cdot d\vec{\boldsymbol{r}} \tag{13.40}$$

for any vector $\vec{\boldsymbol{v}}$.

Direct computation shows that $v \wedge w$ looks quite a bit like $\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}$. However, $v \wedge w$ is a 2-form, whereas $\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}$ is again a vector. But the space of 2-forms is also three-dimensional, so we can identify it, too, with ordinary vectors. However, care now needs to be taken to identify as follows:

$$\hat{\boldsymbol{x}} \longleftrightarrow dy \wedge dz, \tag{13.41}$$

$$\hat{\boldsymbol{y}} \longleftrightarrow dz \wedge dx, \tag{13.42}$$

$$\hat{\boldsymbol{z}} \longleftrightarrow dx \wedge dy. \tag{13.43}$$

This identification also turns vectors into things we can integrate, but in this case using surface integrals, rather than line integrals; we are now mapping $\vec{\boldsymbol{v}}$ to the 2-form $\vec{\boldsymbol{v}} \cdot d\vec{\boldsymbol{A}}$, representing a flux in the $\vec{\boldsymbol{v}}$ direction.

These identifications demonstrate several important properties of the wedge and cross products. First of all, the cross product doesn't really take vectors to vectors! Recall the geometric definition: The magnitude of the cross product is the *area* spanned by the two vectors. Thus, if each vector has the dimensions of length, the cross product has the dimensions of area and is therefore a different object, living in a different vector space.

Another way to see this distinction is to consider a *parity reversal*, or *space inversion*, in which every point in space is taken to its opposite. Thus, each vector $\vec{\boldsymbol{v}}$ is taken to $-\vec{\boldsymbol{v}}$. The cross product of two vectors then transforms with *two* minus signs, and remains unchanged! Historically, such vectors were called *pseudovectors*; see Section 14.12. In the language of differential forms, the "cross product" of two 1-forms is indeed a different object: a 2-form.

This comparison also sheds some light on how the wedge product can be associative, while the cross product is not. Associativity requires the multiplication of *three* objects, and one must keep track of where the various products live.

A by-product of our analysis of the cross product is that we have implicitly established a map between the 1-forms and 2-forms on $\mathbb{R}^3$, as also discussed in Section 12.4. Such identifications between differential forms of different ranks will be discussed further in Section 14.7.

13.6 THE DOT PRODUCT

Consider now the wedge product of the 1-form

$$v = v_x\, dx + v_y\, dy + v_z\, dz \tag{13.44}$$

and the 2-form

$$w = w_x\, dy \wedge dz + w_y\, dz \wedge dx + w_z\, dx \wedge dy. \tag{13.45}$$

The result is a 3-form, and there is only one independent 3-form, so each term is a multiple of every other. Adding them up, we obtain

$$v \wedge w = (v_x w_x + v_y w_y + v_z w_z)\, dx \wedge dy \wedge dz, \tag{13.46}$$

which looks very much like the dot product between $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$! But the dot product is a scalar, so we really want just the coefficient of our basis 3-form $dx \wedge dy \wedge dz$.

As with the cross product, a by-product of our analysis of the dot product is to establish an implicit map, this time between the 0-forms and 3-forms on $\mathbb{R}^3$, as briefly discussed in Section 12.4; such identifications will be discussed further in Section 14.7.

13.7 PRODUCTS OF DIFFERENTIAL FORMS

We can of course multiply any differential forms together, not just 1-forms. So let $\alpha \in \bigwedge^p$ and $\beta \in \bigwedge^q$. Clearly we must have $\alpha \wedge \beta \in \bigwedge^{p+q}$. For example, if

$$\alpha = dx \wedge dy, \tag{13.47}$$

$$\beta = dz \wedge du \wedge dv, \tag{13.48}$$

then

$$\alpha \wedge \beta = dx \wedge dy \wedge dz \wedge du \wedge dv. \tag{13.49}$$

What about $\beta \wedge \alpha$? Direct computation shows that, in this case, $\beta \wedge \alpha = \alpha \wedge \beta$. Why? Because you have to move three 1-forms through two 1-forms, for a total of six sign changes. Thus, in general,

$$\beta \wedge \alpha = (-1)^{pq}\, \alpha \wedge \beta. \tag{13.50}$$

Our original antisymmetry rule for 1-forms is just the special case $p = q = 1$. The simple argument given in Section 11.4 shows that

$$\alpha \wedge \alpha = 0 \tag{13.51}$$

for any p-form with p *odd*, but this argument fails for p *even*. Can you find a p-form α such that $\alpha \wedge \alpha \neq 0$?

13.8 PICTURES OF DIFFERENTIAL FORMS

How do vectors act on functions? By telling you how much the function changes in the direction of the vector. This action is given by

$$\vec{\boldsymbol{v}}(f) = \vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{\nabla}} f \tag{13.52}$$

and involves differentiation. But we have also seen that the 1-form df corresponds to $\vec{\boldsymbol{\nabla}} f$, through the Master Formula (13.4), and we can therefore regard (13.52) as an action of the 1-form df on the vector $\vec{\boldsymbol{v}}$, given by

$$df(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{v}}(f) = \vec{\boldsymbol{\nabla}} f \cdot \vec{\boldsymbol{v}}. \tag{13.53}$$

This action does *not* involve differentiation, only linear algebra, and therefore lends itself to a geometric interpretation in terms of *stacks*.[2]

Consider first the 1-form dx. From calculus, we know that $dx = 0$ along surfaces of the form $x =$ constant. What does $dx(\vec{\boldsymbol{v}})$ represent? We have $dx = \vec{\boldsymbol{\nabla}} x \cdot d\vec{\boldsymbol{r}} = \hat{\boldsymbol{x}} \cdot d\vec{\boldsymbol{r}}$, so $dx(\vec{\boldsymbol{v}}) = \hat{\boldsymbol{x}} \cdot \vec{\boldsymbol{v}} = v_x$ is just the x-component of $\vec{\boldsymbol{v}}$. How can we display this information visually?

We represent dx as an *oriented stack* of surfaces $x =$ constant. Since we have associated dx with $\hat{\boldsymbol{x}}$, we choose unit spacing between the surfaces. Thus, dx corresponds to the drawing in Figure 13.1, where the arrow gives the orientation, namely the direction in which x is increasing. To evaluate $dx(\vec{\boldsymbol{v}})$, all we need to do is superimpose $\vec{\boldsymbol{v}}$ on this stack, as shown in Figure 13.2, then count how many surfaces $\vec{\boldsymbol{v}}$ goes through, which in this case is $\frac{3}{2}$. If the stack had been oriented the other way (corresponding to $-dx$), then the answer would have been negative.

Now that we know how to represent dx, how can we represent $2\,dx$? Clearly, the result of acting on a vector with $2\,dx$ should be twice that of acting on it with dx. But the vector doesn't change, so it must go through more surfaces. Thus, $2\,dx$ is represented by the same drawing as in Figure 13.1, but with the spacing between the surfaces *reduced* by a factor of two, as shown in Figure 13.3. This construction should remind you of a topographic map, where the contour lines are closer together the steeper the terrain is.

[2] A good discussion of *stacks* and related structures in three dimensions can be found in Weinreich [12].

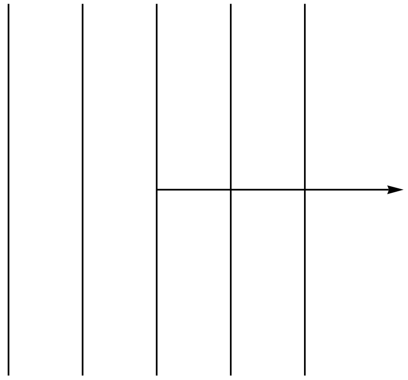


FIGURE 13.1. The stack corresponding to dx.

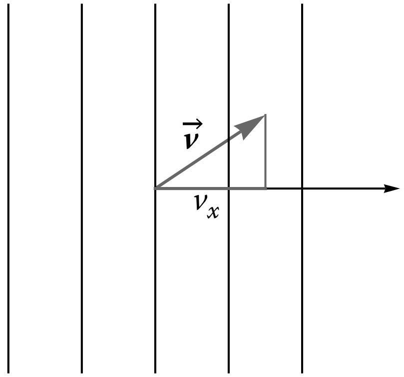


FIGURE 13.2. The geometric evaluation of $dx(\vec{\boldsymbol{v}}) = v_x$.

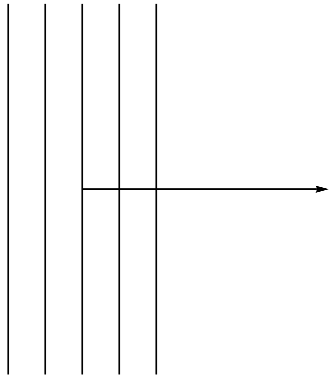


FIGURE 13.3. The stack corresponding to $2\,dx$.

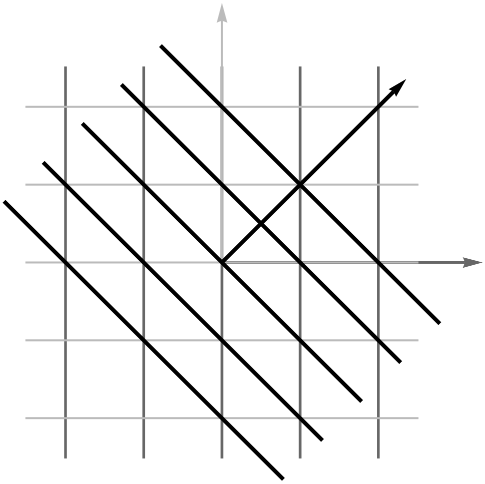


FIGURE 13.4. Adding the stacks dx and dy to get the stack $dx + dy$.

How do we add stacks? Remarkably, the process is simple: just superimpose the corresponding stacks, and connect points of intersection. This construction is shown in Figure 13.4 for the addition of dx and dy. It is straightforward but somewhat tedious to verify that this procedure yields the correct spacing between surfaces, which is $\frac{1}{\sqrt{a^2+b^2}}$ for $a\,dx + b\,dy$. (Why?)

The description of stacks given above works fine for *constant* 1-forms, but not for 1-form *fields*. Just as a vector field is a vector at each point, a 1-form field is a 1-form at each point. Thus, to represent a 1-form field, we need a stack at each point.

For example, Figure 13.5 shows the stacks corresponding to $r\,dr = x\,dx + y\,dy$, and Figure 13.6 shows the stacks corresponding to $r^2\,d\phi = -y\,dx + x\,dy$. Since $r\,dr = \frac{1}{2}\,d(r^2)$, the first example could have been represented by a topographic map whose contours are circles; this 1-form corresponds to the *conservative* vector field $r\,\hat{\boldsymbol{r}} = x\,\hat{\boldsymbol{x}} + y\,\hat{\boldsymbol{y}}$. However, the vector field $-y\,\hat{\boldsymbol{x}} + x\,\hat{\boldsymbol{y}}$ is *not* conservative; the second example does *not* correspond to any topographic map. Stacks provide a useful geometric representation in both cases, but they do not in general combine to form a global representation in terms of contours.[3]

[3]For this reason, some mathematicians regard such pictures as an oversimplification; see Bachman [13].

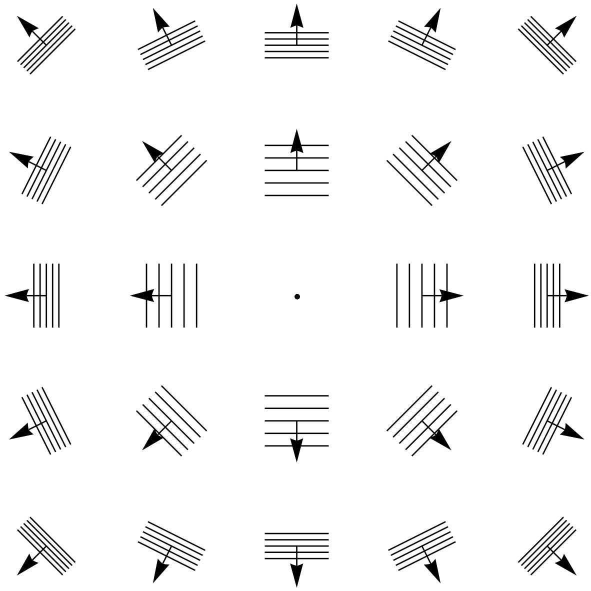


FIGURE 13.5. The stacks corresponding to $r\,dr = x\,dx + y\,dy$.

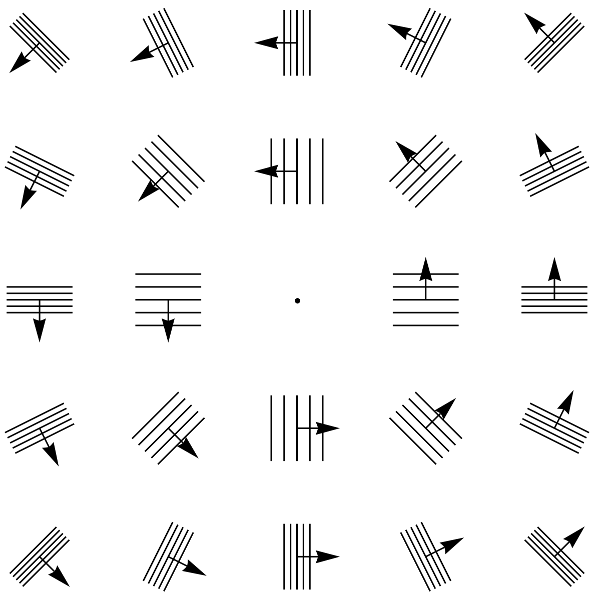


FIGURE 13.6. The stacks corresponding to $r^2\,d\phi = -y\,dx + x\,dy$.

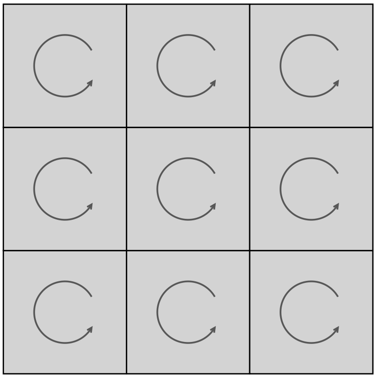


FIGURE 13.7. A representation of $dx \wedge dy$.

Similar methods can be used to represent higher-order forms. For example, the 2-form $dx \wedge dy$ is shown in Figure 13.7. The 2-form $dx \wedge dy$ acts on *pairs* of vectors $(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}})$. How? Construct the parallelogram whose sides are $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$, then superimpose it on the drawing; the number of "cells" covered by the parallelogram gives the result of $dx \wedge dy$ acting on $(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}})$, with the overall sign determined by the relative orientations. Again, this drawing shows a *constant* 2-form; a general 2-form (field) would be represented by a similar drawing at each point.

Similar constructions apply in higher dimensions.[4]

13.9 TENSORS

We now have two quite different interpretations of df:

- df represents a small change in f (infinitesimals);
- $df(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{\nabla}} f \cdot \vec{\boldsymbol{v}}$ (linear map on vectors).

The first interpretation comes from calculus; df is a *differential*. The second comes from linear algebra; df is a *tensor*.[5] We will use both of these interpretations: The calculus of differentials allows us to relate differentials

[4] You can find more pictures of differential forms in Chapter 4 of Misner, Thorne, and Wheeler [14].

[5] More generally, a *tensor* is any multilinear map on vectors. From this point of view, all differential forms are tensors; p-forms are the *antisymmetric* multilinear maps on p vectors.

of different functions, and tensor algebra tells us how to pair vectors and 1-forms (so that, for example, dx and $\hat{\boldsymbol{x}}$ contain the same physical information). Both interpretations contribute to the pictures in Section 13.8.

You may wonder how dx can be small (an infinitesimal) and large (equivalent to a unit vector) at the same time. The answer comes from studying Figures 13.5 and 13.6. Calculus is about *linearization*, the process of zooming in so that functions become linear. The 1-forms shown in these figures are represented by a stack at each point. They are small, because they live at a single point, which is as zoomed in as possible. But they are also large, because *at that point* they behave like unit vectors.

Differential geometers often carry this analogy even further. We introduced the action (13.52) of a vector on a function, which is just the directional derivative of f along $\vec{\boldsymbol{v}}$. In particular, we have

$$\hat{\boldsymbol{x}}(f) = \hat{\boldsymbol{x}} \cdot \vec{\boldsymbol{\nabla}} f = \frac{\partial f}{\partial x}, \tag{13.54}$$

which allows us to identify basis vectors such as $\hat{\boldsymbol{x}}$ with partial derivatives in coordinate directions. For this reason, differential geometers not only use $\{dx^i\}$ as a basis for the 1-forms on $\mathbb{R}^n$ but also use $\left\{\frac{\partial}{\partial x^i}\right\}$ as a basis for the vector fields on $\mathbb{R}^n$.

13.10 INNER PRODUCTS

We have argued that 1-forms are "like" vectors, and that in particular dx is "like" $\hat{\boldsymbol{x}}$. Since $\{\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}\}$ is an orthonormal basis for vectors in $\mathbb{R}^3$, we would like to regard $\{dx, dy, dz\}$ as an orthonormal basis for 1-forms.

We therefore introduce an *inner product* g on $\bigwedge^1(\mathbb{R}^3)$ by *defining* the rectangular basis $\{dx, dy, dz\}$ to be orthonormal under this product, that is,

$$g(dx, dx) = g(dy, dy) = g(dz, dz) = 1, \tag{13.55}$$

$$g(dx, dy) = g(dy, dz) = g(dz, dx) = 0. \tag{13.56}$$

To be an inner product, g must be *linear*, that is,

$$g(f\alpha + \beta, \gamma) = f g(\alpha, \gamma) + g(\beta, \gamma); \tag{13.57}$$

symmetric, that is,

$$g(\beta, \alpha) = g(\alpha, \beta); \tag{13.58}$$

and *nondegenerate*, that is,

$$g(\alpha, \beta) = 0 \ \forall \beta \Longrightarrow \alpha = 0, \tag{13.59}$$

and it is easily checked that (13.57)–(13.59) follow from the corresponding properties of the dot product. However, we do *not* require that g be *positive definite*, which would force $g(\alpha,\alpha) \geq 0$, with $g(\alpha,\alpha) = 0$ if and only if $\alpha = 0$. As we will see, this property does not hold in either special or general relativity.

13.11 POLAR COORDINATES II

As an example, consider polar coordinates. We have

$$r^2 = x^2 + y^2, \tag{13.60}$$

$$\tan\phi = \frac{y}{x} \tag{13.61}$$

so that

$$r\,dr = x\,dx + y\,dy, \tag{13.62}$$

$$\frac{r^2}{x^2}\,d\phi = (1 + \tan^2\phi)\,d\phi = \frac{x\,dy - y\,dx}{x^2}. \tag{13.63}$$

Thus,

$$g(r\,dr, r\,dr) = g(x\,dx + y\,dy, x\,dx + y\,dy) = x^2 + y^2 = r^2, \tag{13.64}$$

$$g(r^2\,d\phi, r^2\,d\phi) = g(x\,dy - y\,dx, x\,dy - y\,dx) = x^2 + y^2 = r^2, \tag{13.65}$$

$$g(r\,dr, r^2\,d\phi) = g(x\,dx + y\,dy, x\,dy - y\,dx) = yx - xy = 0, \tag{13.66}$$

and we conclude that $\{dr, r\,d\phi\}$ is also an orthonormal basis.

This is not surprising! Recall that infinitesimal arclength is given by

$$ds^2 = dx^2 + dy^2 = dr^2 + r^2\,d\phi^2. \tag{13.67}$$

Thus, the line element tells us directly which basis of 1-forms is orthonormal.

We can take this analogy even further by considering the vector differential

$$d\vec{\boldsymbol{r}} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}}, \tag{13.68}$$

which we now interpret as a *vector-valued 1-form*. As usual, we have

$$d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = ds^2 \tag{13.69}$$

so we can think of $d\vec{\boldsymbol{r}}$ as the "square root" of ds^2. Again, $d\vec{\boldsymbol{r}}$ tells us directly which 1-forms are orthonormal.

We will develop these ideas further in Chapter 14.

≻ CHAPTER 14 ≺

HODGE DUALITY

14.1 BASES FOR DIFFERENTIAL FORMS

Let M be an n-dimensional surface, with coordinates (x^i). You can think of M as being $\mathbb{R}^n$, but we will also consider surfaces in higher-dimensional spaces, such as the 2-sphere in $\mathbb{R}^3$. Then, as before, $\bigwedge^1(M)$ is the span of the 1-forms $\{dx^i\}$, $\bigwedge^0(M)$ is the space of functions on M, and $\bigwedge^p(M)$ is the space of p-forms; we will often write simply $\bigwedge^p$ for $\bigwedge^p(M)$.

The basis $\{dx^i\}$ of 1-forms is quite natural and is called a *coordinate basis*. However, in the presence of an inner product g on $\bigwedge^1$, it is often preferable to work instead with an orthonormal basis. So let's start over and work with an arbitrary (for now) basis. Suppose $\{\sigma^i\}$ is a basis of $\bigwedge^1$. Then a basis for $\bigwedge^p$ is

$$\{\sigma^I\} = \{\sigma^{i_1} \wedge ... \wedge \sigma^{i_p}\}, \tag{14.1}$$

where the *index set* $I = \{i_1, ..., i_p\}$ satisfies $1 \leq i_1 < ... < i_p \leq n$, and where we are of course assuming $p \leq n$.

What about $\bigwedge^0$, the space of scalars, that is, functions? The elements of this space contain no factors of σ^i. There is nonetheless a natural basis for this one-dimensional space, namely the constant function 1.

14.2 THE METRIC TENSOR

Now suppose there is an inner product g on $\bigwedge^1$. The *components* of g in the basis $\{\sigma^i\}$ are given by

$$g^{ij} = g(\sigma^i, \sigma^j). \tag{14.2}$$

As usual, we will also use g to denote the matrix (g^{ij}).

Recall that an inner product is symmetric; thus, $g^{ji} = g^{ij}$, so the matrix g is symmetric. An inner product must also be non-degenerate, and a little thought shows that this is equivalent to requiring that the determinant $|g|$

be nonzero; g must be invertible. The inner product g is often called the *metric tensor*, or, since it acts on 1-forms (and not vectors), the *inverse metric*.

But any symmetric matrix can be diagonalized, and if the determinant is nonzero, then none of the diagonal entries can be 0. Equivalently, we can assume that the basis $\{\sigma^i\}$ is *orthogonal*:

$$g(\sigma^i, \sigma^j) = 0 \qquad (i \neq j). \tag{14.3}$$

Furthermore, the basis can be rescaled, so that

$$|g(\sigma^i, \sigma^i)| = 1. \tag{14.4}$$

However, since we are not assuming that g is positive definite, we cannot determine the signs. Putting this together, we can assume without loss of generality that our basis is *orthonormal*, that is, that

$$g(\sigma^i, \sigma^j) = \pm\delta^{ij}, \tag{14.5}$$

where δ^{ij} is the Kronecker delta, which is 1 if $i = j$, and 0 otherwise.

The information in g can be encoded in other ways. A standard description is to give the *line element*, which, in a coordinate basis, is

$$ds^2 = g_{ij}\, dx^i dx^j, \tag{14.6}$$

where the matrix (g_{ij}) turns out to be the inverse of the matrix (g^{ij}) (for the basis $\{\sigma^i = dx^i\}$). If our basis is orthonormal, then (the matrix) g is its own inverse!

An example we have already seen is polar coordinates, for which

$$ds^2 = dr^2 + r^2\, d\phi^2, \tag{14.7}$$

which tells us that $\{dr, r\, d\phi\}$ is an orthonormal basis (and $\{dr, d\phi\}$ is a coordinate basis). From the point of view of differential forms, the line element is nothing more than a way of encoding the metric.

Equivalent information is also contained in the vector differential $d\vec{\boldsymbol{r}}$, which is most useful in an *orthogonal* coordinate system, that is, when

$$g(dx^i, dx^j) = 0 \qquad (i \neq j). \tag{14.8}$$

In this case, each σ^i is obtained by normalizing dx^i, and we write

$$d\vec{\boldsymbol{r}} = \sigma^i \hat{e}_i, \tag{14.9}$$

where $\hat{e}_i$ is the unit vector in the x^i direction. Thus,

$$d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = ds^2 \tag{14.10}$$

as expected. From the point of view of differential forms, $d\vec{\boldsymbol{r}}$ is just a *vector-valued 1-form.*

Again, we have seen this before in polar coordinates, where

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + (r\,d\phi)\,\hat{\boldsymbol{\phi}}. \tag{14.11}$$

14.3 SIGNATURE

Although we do not know whether each $g(\sigma^i, \sigma^i)$ is $+1$ or -1, it is easily seen that the number of plus signs (p) and minus signs (m) is basis independent. We define the *signature* of the metric g to be

$$s = m, \tag{14.12}$$

and therefore the signature is just the number of minus signs.[1]

Of particular interest are the cases $s = 0$, for which g is positive definite, which gives rise to *Riemannian geometry*, and $s = 1$, for which there is precisely one minus sign, which is called *Lorentzian geometry*, and which is the geometric arena for (both special and general) relativity.

Particular examples are *Euclidean 2-space*, usually denoted $\mathbb{R}^2$ (rather than the more accurate $\mathbb{E}^2$), with line element

$$ds^2 = dx^2 + dy^2, \tag{14.13}$$

and *Minkowski 2-space*, $\mathbb{M}^2$, with line element

$$ds^2 = dx^2 - dt^2. \tag{14.14}$$

14.4 INNER PRODUCTS OF HIGHER-RANK FORMS

When comparing the cross product in $\mathbb{R}^3$ with the wedge product on $\bigwedge^1(\mathbb{R}^3)$, it was natural to associate $\hat{\boldsymbol{z}}$ with *both* $dz \in \bigwedge^1$ and $dx \wedge dy \in \bigwedge^2$. Since $|\hat{\boldsymbol{z}}| = 1$, we expect not only

$$g(dz, dz) = 1, \tag{14.15}$$

[1]Many authors instead define the signature to be $p - m$, the difference between the number of plus and minus signs. Our choice has the advantage that the metric in relativity always has the same signature.

which we already know, but also something like

$$g(dx \wedge dy, dx \wedge dy) = 1, \tag{14.16}$$

which we have yet to define. Consider for simplicity $\bigwedge^2(\mathbb{R}^2)$, which only has one independent element. Suppose we *define* a linear operator, also called g, on $\bigwedge^2$ by (14.16). What properties does it have? Well,

$$\begin{aligned} g(\alpha \wedge \beta, dx \wedge dy) &= g\big((\alpha_x\, dx + \alpha_y\, dy) \wedge (\beta_x\, dx + \beta_y\, dy), dx \wedge dy)\big) \\ &= g\big((\alpha_x\beta_y - \alpha_y\beta_x)\, dx \wedge dy, dx \wedge dy\big) \\ &= (\alpha_x\beta_y - \alpha_y\beta_x) g(dx \wedge dy, dx \wedge dy) \\ &= \alpha_x\beta_y - \alpha_y\beta_x \\ &= \begin{vmatrix} g(\alpha, dx) & g(\alpha, dy) \\ g(\beta, dx) & g(\beta, dy) \end{vmatrix}. \end{aligned} \tag{14.17}$$

It seems reasonable to conjecture that

$$g(\alpha \wedge \beta, \gamma \wedge \delta) = \begin{vmatrix} g(\alpha, \gamma) & g(\alpha, \delta) \\ g(\beta, \gamma) & g(\beta, \delta) \end{vmatrix}. \tag{14.18}$$

So let's start again and take (14.18) as the *definition* of the metric g acting on 2-forms. A similar $p \times p$ determinant can be used to define g on $\bigwedge^p$. It is straightforward to check that in an orthonormal basis $\{\sigma^i\}$, not only is

$$g(\sigma^i, \sigma^j) = \pm\delta^{ij} \tag{14.19}$$

but also

$$g(\sigma^I, \sigma^J) = \pm\delta^{IJ} \tag{14.20}$$

so that our canonical basis $\{\sigma^I\}$ of $\bigwedge^p$ is orthonormal as well—precisely what we wanted.

What about the special case $\bigwedge^0$, the space of scalars? We chose the constant function 1 as the basis of this one-dimensional space and would like this basis to be normalized. Thus, we set

$$g(1, 1) = 1 \tag{14.21}$$

so that $g(f_1, f_2) = f_1 f_2$ for any functions f_1 and f_2.

14.5 THE SCHWARZ INEQUALITY

We have just seen that if $\alpha, \beta \in \bigwedge^1$, then

$$\begin{aligned} g(\alpha\wedge\beta,\alpha\wedge\beta) &= \begin{vmatrix} g(\alpha,\alpha) & g(\alpha,\beta) \\ g(\beta,\alpha) & g(\beta,\beta) \end{vmatrix} \\ &= g(\alpha,\alpha)g(\beta,\beta) - g(\alpha,\beta)^2. \end{aligned} \tag{14.22}$$

But we also know that over $\mathbb{R}^2$ we must have

$$\alpha\wedge\beta = f\,dx\wedge dy \tag{14.23}$$

for some function f (which was computed explicitly in (14.17)), so that[2]

$$\begin{aligned} g(\alpha\wedge\beta,\alpha\wedge\beta) &= f^2\,g(dx\wedge dy, dx\wedge dy) \\ &= f^2\,g(dx,dx)\,g(dy,dy) \\ &= f^2 \geq 0. \end{aligned} \tag{14.24}$$

Putting these two computations together, we obtain the *Schwarz inequality*, which says that, in Euclidean space,

$$g(\alpha,\beta)^2 \leq g(\alpha,\alpha)\,g(\beta,\beta). \tag{14.25}$$

Consider the same argument in (two-dimensional) Minkowski space. The first equation is unchanged, but we now have

$$\alpha\wedge\beta = f\,dx\wedge dt \tag{14.26}$$

and the last computation becomes

$$\begin{aligned} g(\alpha\wedge\beta,\alpha\wedge\beta) &= f^2\,g(dx\wedge dt, dx\wedge dt) \\ &= f^2\,g(dx,dx)\,g(dt,dt) \\ &= -f^2 \leq 0, \end{aligned} \tag{14.27}$$

which yields the *reverse Schwarz inequality*, namely

$$g(\alpha,\beta)^2 \geq g(\alpha,\alpha)\,g(\beta,\beta). \tag{14.28}$$

[2]In Euclidean signature, this result must hold in any number of dimensions, since α, β span a two-dimensional, Euclidean subalgebra, in which the computation takes place as outlined.

14.6 ORIENTATION

As before, let $\{\sigma^i\}$ be an orthonormal basis of $\bigwedge^1(M)$, and consider $\bigwedge^n(M)$. First of all, this space is one-dimensional, with basis

$$\omega = \sigma^1 \wedge \ldots \wedge \sigma^n. \tag{14.29}$$

But ω is a unit n-form, since

$$g(\omega, \omega) = g(\sigma^1, \sigma^1)...g(\sigma^n, \sigma^n) = (-1)^s. \tag{14.30}$$

Furthermore, there are precisely two unit n-forms, namely $\pm\omega$, a statement that does not, of course, depend on the choice of basis.

An *orientation* ω on M is a choice of one of the two unit n-forms. If an orthonormal basis is given explicitly, we will always assume that ω is precisely as given in (14.29), that is, that the basis is ordered to match (or define) the orientation.

14.7 THE HODGE DUAL

We have seen that, in $\mathbb{R}^3$, the wedge product of two 1-forms looks very much like the cross product of vectors, except that it is a 2-form. And the wedge product of a 1-form and a 2-form looks very much like the dot product of vectors, except that it is a 3-form. There is clearly some sort of correspondence of the form

$$dx \longleftrightarrow dy \wedge dz, \tag{14.31}$$

$$dy \longleftrightarrow dz \wedge dx, \tag{14.32}$$

$$dz \longleftrightarrow dx \wedge dy, \tag{14.33}$$

$$1 \longleftrightarrow dx \wedge dy \wedge dz. \tag{14.34}$$

This correspondence is the *Hodge dual* operation, which we now describe.

What do the correspondences (14.31)–(14.34) have in common? For each of the eight basis forms that appear therein, call it α, the corresponding form, call it $*\alpha$, satisfies

$$\alpha \wedge *\alpha = \omega. \tag{14.35}$$

This is almost enough to define the Hodge dual operation, denoted by $*$. Each of the above basis elements α is a unit p-form. What happens if we

rescale α? In order for the Hodge dual to be a linear map, $*\alpha$ must be rescaled as well, so we must have

$$\alpha \wedge *\alpha = g(\alpha, \alpha)\, \omega, \tag{14.36}$$

and this is indeed a defining property of the Hodge dual.[3]

We are not quite done. Comparing $(\alpha + \beta) \wedge *(\alpha + \beta)$ with $\alpha \wedge *\alpha$ and $\beta \wedge *\beta$ yields

$$\alpha \wedge *\beta + \beta \wedge *\alpha = 2g(\alpha, \beta)\, \omega, \tag{14.37}$$

a technique known as *polarization*. We would like the left-hand side to be symmetric,[4] so we are finally led to define the Hodge dual operator $*$ by

$$\alpha \wedge *\beta = g(\alpha, \beta)\, \omega. \tag{14.38}$$

In using this definition, it is convenient to extend the notation so as to include 0-forms. We have already defined g on 0-forms by (14.21) and wedge products involving 0-forms by

$$f \wedge \alpha = f\, \alpha \tag{14.39}$$

for any form α and function f. It is then straightforward to show that

$$*1 = 1 \wedge *1 = g(1, 1)\, \omega = \omega \tag{14.40}$$

in any signature and dimension. Similarly, from

$$\omega \wedge *\omega = g(\omega, \omega)\, \omega = (-1)^s \omega, \tag{14.41}$$

we have

$$*\omega = (-1)^s. \tag{14.42}$$

The implicit formula (14.38) above in fact suffices to compute the Hodge dual, and it is of course enough to do so on a basis. Since (14.38) must hold for *all* forms, each orthonormal basis element of the form $\sigma^{i_1} \wedge ... \wedge \sigma^{i_p}$ in $\bigwedge^p$ must be mapped to its "complement" in $\bigwedge^{n-p}$ of the form $\pm\sigma^{i_{p+1}} \wedge ... \wedge \sigma^{i_n}$, where (i_p) is a permutation of $(1...n)$; it only remains to determine the signs.

[3] Many authors use instead the convention

$$\alpha \wedge *\alpha = g(*\alpha, *\alpha)\, \omega,$$

which turns out to differ from (14.38) by a factor of $(-1)^s$. In the positive-definite case, which is all that many authors consider, $s = 0$ and the two conventions agree.

[4] One must check that our definition actually has this property; see Section 14.14.

14.8 HODGE DUAL IN MINKOWSKI 2-SPACE

Consider Minkowski 2-space, with line element

$$ds^2 = dx^2 - dt^2 \tag{14.43}$$

and (ordered) orthonormal basis $\{dx, dt\}$. The orientation is

$$\omega = dx \wedge dt, \tag{14.44}$$

and it is straightforward to compute the Hodge dual on a basis. First of all, we have

$$dx \wedge *dx = g(dx, dx)\, dx \wedge dt = dx \wedge dt, \tag{14.45}$$

from which it follows that[5]

$$*dx = dt. \tag{14.46}$$

Similarly, from

$$dt \wedge *dt = g(dt, dt)\, dx \wedge dt = -dx \wedge dt, \tag{14.47}$$

we see that

$$*dt = dx. \tag{14.48}$$

The remaining cases were already worked out in general; for Minkowski 2-space we obtain

$$*1 = dx \wedge dt, \tag{14.49}$$

$$*(dx \wedge dt) = -1. \tag{14.50}$$

14.9 HODGE DUAL IN EUCLIDEAN 2-SPACE

Consider Euclidean 2-space, with line element

$$ds^2 = dx^2 + dy^2 \tag{14.51}$$

and (ordered) orthonormal basis $\{dx, dy\}$. The orientation is

$$\omega = dx \wedge dy, \tag{14.52}$$

[5]One must also use the fact that $dt \wedge *dx = 0$ to show that there is no dx term in $*dx$. This argument generalizes to other, similar computations and will not be repeated.

and it is again straightforward to compute the Hodge dual on a basis. In analogy with Section 14.8, we have

$$dx \wedge *dx = g(dx, dx)\, dx \wedge dy = dx \wedge dy, \tag{14.53}$$

from which it follows that

$$*dx = dy. \tag{14.54}$$

Similarly, from

$$dy \wedge *dy = g(dy, dy)\, dx \wedge dy = dx \wedge dy, \tag{14.55}$$

we see that

$$*dy = -dx. \tag{14.56}$$

As before, the remaining cases were already worked out in general; for Euclidean 2-space we obtain

$$*1 = dx \wedge dy, \tag{14.57}$$

$$*(dx \wedge dy) = 1. \tag{14.58}$$

It is important to note that, even in this positive-definite setting, there are some factors of -1 in these relations, which arise from the necessary permutations. You should verify for yourself that no such factors arise in Euclidean 3-space.

14.10 HODGE DUAL IN POLAR COORDINATES

Consider polar coordinates, with line element

$$ds^2 = dr^2 + r^2\, d\phi^2 \tag{14.59}$$

and (ordered) orthonormal basis $\{dr, r\, d\phi\}$. The orientation is

$$\omega = dr \wedge r\, d\phi, \tag{14.60}$$

and the computation is nearly identical to the one in Section 14.9. We have

$$dr \wedge *dr = g(dr, dr)\, dr \wedge r\, d\phi = dr \wedge r\, d\phi, \tag{14.61}$$

from which it follows that

$$*dr = r\, d\phi. \tag{14.62}$$

Similarly, from

$$r\,d\phi \wedge *r\,d\phi = g(r\,d\phi, r\,d\phi)\,dr \wedge r\,d\phi = dr \wedge r\,d\phi, \tag{14.63}$$

we see that

$$*r\,d\phi = -dr. \tag{14.64}$$

What about $*d\phi$? The Hodge dual is linear, so

$$*d\phi = *\left(\frac{r}{r}\,d\phi\right) = \frac{1}{r}*(r\,d\phi) = -\frac{1}{r}dr. \tag{14.65}$$

Be careful when working with differential forms that are not normalized; it is often easiest to first normalize them, compute the dual of the normalized form, and then undo the normalization using linearity.

14.11 DOT AND CROSS PRODUCT REVISITED

We can now finally write down precise equivalents of the dot and cross products using differential forms. First of all, the definition of the Hodge dual tells us that the Hodge dual, $*$, and the metric, g, in fact contain the same information. Interpreting g on 1-forms as the dot product, we have

$$\alpha \cdot \beta = (-1)^s *(\alpha \wedge *\beta), \tag{14.66}$$

where we have used (14.38) and (14.42). The dot product is thus defined in any dimension and signature, and reduces to the expected dot product in the positive-definite case. In three dimensions, if we identify $\hat{\boldsymbol{x}}$ with dx, etc., we recover the dot product in its usual form.

As for the cross product, recall that $\alpha \wedge \beta$ over $\bigwedge(\mathbb{R}^3)$ had the form of the cross product but was a 2-form. We can now use the Hodge dual to turn this into a 1-form, obtaining

$$\alpha \times \beta = *(\alpha \wedge \beta). \tag{14.67}$$

Note, however, that this only works in three dimensions; otherwise the result is not a 1-form.[6]

[6] A generalization of the cross product with many of the same properties can be obtained as the product of $n-1$ 1-forms in n dimensions given by

$$\alpha_1 \times \ldots \times \alpha_{n-1} = *(\alpha_1 \wedge \ldots \wedge \alpha_{n-1}).$$

However, a *bilinear* map with the properties of the cross product is only possible in three and, surprisingly, in seven dimensions. (The latter product is related to the octonion multiplication table, just as the ordinary three-dimensional cross product is related to the quaternion multiplication table.)

14.12 PSEUDOVECTORS AND PSEUDOSCALARS

Is the cross product of two vectors a vector? Yes and no. Yes, of course, $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ is *a* vector, but is it the same type of vector as $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$? Not really.

First of all, the dimensions are different: If $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ have dimensions of length, then $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ has dimensions of area. This behavior arises from the geometric definition of the cross product as *directed area*.

A more interesting difficulty arises when considering symmetries. What happens if one reflects points through the origin, a so-called *parity* transformation? Each vector $\vec{\boldsymbol{u}}$ would transform into $-\vec{\boldsymbol{u}}$. But $\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$ would transform into itself, since $(-\vec{\boldsymbol{u}}) \times (-\vec{\boldsymbol{v}}) = \vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}$!

As already discussed in Section 13.5, there thus appear to be two types of vectors, which were historically called *vectors* and *pseudovectors*, respectively. Similarly, there are two types of scalars, called *pseudoscalars* and *scalars*, depending on whether they do, respectively do not, change sign under a parity transformation.

Now recall that over (Euclidean) $\mathbb{R}^3$ we have *two* three-dimensional spaces of forms (the 1- and 2-forms), and *two* one-dimensional spaces (the 0- and 3-forms). How do forms behave under a parity transformation? Each basis 1-form transforms into its negative, so odd-rank forms pick up a minus sign, whereas even-rank forms remain the same. With the wisdom of hindsight, it is now easy to identify not only 0-forms as scalars and 1-forms as vectors, but also 2-forms as pseudovectors and 3-forms as pseudoscalars.

14.13 THE GENERAL CASE

We now compute the Hodge dual on an orthonormal basis directly from the definition. Recall that

$$\alpha \wedge *\beta = g(\alpha, \beta)\, \omega. \tag{14.68}$$

Suppose σ^I is a basis p-form in $\bigwedge^p$. By permuting the basis 1-forms σ^i, we can bring σ^I to the form

$$\sigma^I = \sigma^1 \wedge ... \wedge \sigma^p. \tag{14.69}$$

Furthermore, we can assume without loss of generality that our permutation is *even*, and hence does not change the orientation ω.[7] Consider now the defining property

$$\sigma^I \wedge *\sigma^I = g(\sigma^I, \sigma^I)\, \omega, \tag{14.70}$$

[7] The only tricky case is if $p = n-1$, when it may be necessary to replace σ^I with $-\sigma^I$.

from which it is apparent that $*\sigma^I$ must contain a term of the form

$$\sigma^J = \sigma^{p+1} \wedge \ldots \wedge \sigma^n. \tag{14.71}$$

Furthermore, it is easy to see that $*\sigma^I$ cannot contain any other terms, as otherwise it would not yield zero when wedged with all other basis p-forms, as required by the definition (since our basis is orthonormal). We conclude that

$$*\sigma^I = g(\sigma^I, \sigma^I)\, \sigma^J = g(\sigma^1, \sigma^1) \ldots g(\sigma^p, \sigma^p)\, \sigma^J. \tag{14.72}$$

We can repeat this argument to determine $*\sigma^J$. We have

$$\sigma^J \wedge \sigma^I = (-1)^{p(n-p)} \sigma^I \wedge \sigma^J = (-1)^{p(n-p)} \omega, \tag{14.73}$$

from which it follows that

$$*\sigma^J = (-1)^{p(n-p)} g(\sigma^J, \sigma^J)\, \sigma^I. \tag{14.74}$$

Putting these duals together, we have

$$\begin{aligned} **\sigma^I &= g(\sigma^I, \sigma^I)\, *\sigma^J \\ &= (-1)^{p(n-p)} g(\sigma^I, \sigma^I) g(\sigma^J, \sigma^J)\, \sigma^I \\ &= (-1)^{p(n-p)} g(\sigma^1, \sigma^1) \ldots g(\sigma^n, \sigma^n)\, \sigma^I \\ &= (-1)^{p(n-p)} g(\omega, \omega)\, \sigma^I \\ &= (-1)^{p(n-p)+s}\, \sigma^I. \end{aligned} \tag{14.75}$$

This identity is true for any differential form and is most easily remembered as

$$** = (-1)^{p(n-p)+s}. \tag{14.76}$$

A special case is $s = 0$ and $n = 3$, for which the exponent is always even, so that

$$** = 1, \tag{14.77}$$

in agreement with Section 12.4.

14.14 TECHNICAL NOTE ON THE HODGE DUAL

We really ought to verify that our implicit definition of the Hodge dual, namely

$$\alpha \wedge *\beta = g(\alpha, \beta)\, \omega, \tag{14.78}$$

is in fact well-defined. To do so, we rely on a standard result about linear maps.

Lemma: *Given any linear map $f : V \longmapsto \mathbb{R}$ on a vector space V with (nondegenerate) inner product g, there is a unique element $\gamma \in V$ such that*

$$f(\alpha) = g(\alpha, \gamma) \tag{14.79}$$

for every $\alpha \in V$.

The proof of this lemma is obtained by introducing an orthonormal basis $\{\sigma^i\}$, after which one sees that one must have

$$\gamma = \sum_i \frac{f(\sigma^i)}{g(\sigma^i, \sigma^i)}\, \sigma^i. \tag{14.80}$$

Since the map

$$\alpha \longmapsto \alpha \wedge \gamma \tag{14.81}$$

for $\alpha \in \bigwedge^p$ and fixed $\gamma \in \bigwedge^{n-p}$ is linear (and yields a multiple of the orientation, ω), we can use the lemma to infer the existence of a unique element $h(\gamma)$ such that

$$\alpha \wedge \gamma = g\big(\alpha, h(\gamma)\big)\, \omega. \tag{14.82}$$

We can repeat this construction for arbitrary γ, and it is easily seen that the map $\gamma \longmapsto h(\gamma)$ is itself linear. Furthermore, it is straightforward to evaluate h on our basis. Setting

$$h(\sigma^J) = a_j\, \sigma^{I_j}, \tag{14.83}$$

where $\{\sigma^{I_m}\}$ is an orthonormal basis of p-forms, and σ^J is an $(n-p)$-form, we have

$$\sigma^{I_m} \wedge \sigma^J = g(\sigma^{I_m}, a_j\, \sigma^{I_j})\, \omega = \pm a_m\, \omega. \tag{14.84}$$

Rather than work out the correct sign, it is enough to note that if $J \cap I_m$ is not empty, then $a_m = 0$, whereas if the intersection is empty, then $a_m = \pm 1$. Thus, $h(\sigma^J)$ must be precisely $\pm\sigma^K$, where K is the dual index set to J, so that in particular $\sigma^J \wedge \sigma^K = \pm\omega$.

We are almost there. We have demonstrated the existence of a map h from p-forms to $(n-p)$-forms which is almost, but not quite, the Hodge dual. Regardless of signs, the computation above also demonstrates that this map is invertible—and the inverse is the Hodge dual, that is,

$$h(*\beta) = \beta \tag{14.85}$$

for $\beta \in \bigwedge^p$. With the wisdom of hindsight, we see that

$$h(\gamma) = (-1)^{p(n-p)+s} *\gamma \tag{14.86}$$

or equivalently, replacing p by $n-p$ (which doesn't change anything),

$$*\beta = (-1)^{p(n-p)+s}\, h(\beta). \tag{14.87}$$

Since we have established that h is well-defined, so is $*$.

14.15 APPLICATION: DECOMPOSABLE FORMS

A *decomposable* differential form is one that can be written as a product of 1-forms. This is a nontrivial assertion; linear combinations are *not* allowed.

It is easy to see geometrically that all 2-forms in $\mathbb{R}^3$ must be decomposable. In three dimensions, a 2-form can be thought of as the normal vector to a plane (more generally, a surface), and the wedge product mimics the cross product. But any vector in $\mathbb{R}^3$ can be written as the cross product of two other vectors: choose *any* two linearly independent vectors in the plane orthogonal to the given vector, and rescale appropriately.

To find an example of a 2-form that it *not* decomposable, we need four dimensions, say with coordinates w, x, y, z. Then

$$\alpha = dx \wedge dy + dz \wedge dw \tag{14.88}$$

is not decomposable. Why not? If α were decomposable, then we could write $\alpha = \beta \wedge \gamma$, for some 1-forms β and γ. Clearly $\beta \wedge \gamma \wedge \beta \wedge \gamma = 0$, but direct computation shows that $\alpha \wedge \alpha \neq 0$.

We have in fact proved that $\alpha \wedge \alpha = 0$ for any decomposable form, although the converse is false. Can you find an example of a differential form that is *not* decomposable, yet which "squares" to zero using the wedge product?

We are now able to provide an elegant argument that $(n-1)$-forms in n dimensions are decomposable. Let $\alpha \in \bigwedge^{n-1}$. Then $*\alpha \in \bigwedge^1$ is a 1-form, and we can expand $*\alpha$ to a basis $\{\tau^1 = *\alpha, \tau^2, ..., \tau^n\}$ of $\bigwedge^1$. Since the τ^i are linearly independent,

$$\tau^1 \wedge ... \wedge \tau^n = f\omega \tag{14.89}$$

for some nonzero function f, which further implies that[8]

$$*\tau^1 = h\, \tau^2 \wedge ... \wedge \tau^n \tag{14.90}$$

[8]This argument must be modified if $g(\alpha, \alpha) = 0$.

for some nonzero function h. But now, since $** = \pm 1$,

$$\alpha = \pm *(*\alpha) = \pm *\tau^1 = \pm h\,\tau^2 \wedge \ldots \wedge \tau^n, \tag{14.91}$$

and we're done.

14.16 PROBLEMS

1. **Cross Product**

   Let $\vec{\boldsymbol{u}} = u_1\,\hat{\boldsymbol{x}} + u_2\,\hat{\boldsymbol{y}} + u_3\,\hat{\boldsymbol{z}}$ be an ordinary vector in $\mathbb{R}^3$. Find two vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ such that $\vec{\boldsymbol{u}} = \vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}$.

2. **Decomposable Forms**

   A p-form $\beta \in \bigwedge^p(\mathbb{R}^n)$ is called *decomposable* if there exist 1-forms $\alpha_i \in \bigwedge^1(\mathbb{R}^n)$ with

   $$\beta = \alpha_1 \wedge \ldots \wedge \alpha_p.$$

   (a) Show explicitly that all elements of $\bigwedge^2(\mathbb{R}^3)$, that is, all 2-forms in $\mathbb{R}^3$, are decomposable. In other words, show that

   $$H = H_x\,dy \wedge dz + H_y\,dz \wedge dx + H_z\,dx \wedge dy$$

   is decomposable.

   (b) Show explicitly that all 3-forms are decomposable in $\mathbb{R}^4$.

   (c) Can you extend your construction to show that all elements of $\bigwedge^{n-1}(\mathbb{R}^n)$ are decomposable?

3. **Inner Products I**

   Let $\vec{\boldsymbol{v}} = v^1\,\hat{\boldsymbol{x}} + v^2\,\hat{\boldsymbol{y}}$ and $\vec{\boldsymbol{w}} = w^1\,\hat{\boldsymbol{x}} + w^2\,\hat{\boldsymbol{y}}$ be ordinary vectors in $\mathbb{R}^2$. Consider the inner product defined by

   $$\vec{\boldsymbol{v}} \star \vec{\boldsymbol{w}} = \frac{1}{2}\left(v^1 w^2 + v^2 w^1\right).$$

   (a) Compute $\vec{\boldsymbol{v}} \star \vec{\boldsymbol{v}}$.

   (b) Using only your expression for $\vec{\boldsymbol{v}} \star \vec{\boldsymbol{v}}$, but assuming that it holds for any vector $\vec{\boldsymbol{v}}$ (including $\vec{\boldsymbol{w}}$), derive the expression above for $\vec{\boldsymbol{v}} \star \vec{\boldsymbol{w}}$.

   (c) Find a vector $\vec{\boldsymbol{x}} \in \mathbb{R}^2$ such that $\vec{\boldsymbol{x}} \star \vec{\boldsymbol{x}} = 1$.

   (d) Find a vector $\vec{\boldsymbol{y}} \in \mathbb{R}^2$ such that $\vec{\boldsymbol{y}} \star \vec{\boldsymbol{y}} = -1$.

   (e) How many independent vectors $\vec{\boldsymbol{z}} \in \mathbb{R}^2$ are there satisfying $\vec{\boldsymbol{z}} \star \vec{\boldsymbol{z}} = 0$?

4. **Inner Products II**

   (a) Now work in $\mathbb{R}^3$, and suppose there is an inner product such that
   $$\vec{\boldsymbol{v}} \star \vec{\boldsymbol{v}} = v^1v^2 + v^2v^3 + v^3v^1.$$

   (b) Determine $\vec{\boldsymbol{v}} \star \vec{\boldsymbol{w}}$.

   (c) Find a basis of $\mathbb{R}^3$ such that each basis vector $\vec{\boldsymbol{u}}$ satisfies either $\vec{\boldsymbol{u}} \star \vec{\boldsymbol{u}} = 1$ or $\vec{\boldsymbol{u}} \star \vec{\boldsymbol{u}} = -1$, and any two basis vectors $\vec{\boldsymbol{u}}$, $\vec{\boldsymbol{v}}$ satisfy $\vec{\boldsymbol{u}} \star \vec{\boldsymbol{v}} = 0$.

   (d) Using the above results, or otherwise, rewrite xy as the sum or difference of exactly two squares, and $xy + yz + zx$ as the sum or difference of exactly three squares.

5. **Hodge Dual in Minkowski Space**

   Four-dimensional Minkowski space has an orthonormal, oriented basis of 1-forms given by
   $$\{dx, dy, dz, dt\}$$
   with $g(dt, dt) = -1$, $g(dx, dx) = g(dy, dy) = g(dz, dz) = 1$, and all others zero. The "volume element" (choice of orientation) is given by $\omega = dx \wedge dy \wedge dz \wedge dt$.

   (a) Determine the Hodge dual operator $*$ on all forms by computing its action on basis forms at each rank.

   (b) How does your answer change if the opposite orientation is chosen, namely $\omega = dt \wedge dx \wedge dy \wedge dz$?

6. **Spherical Coordinates**

   Consider spherical coordinates in three-dimensional Euclidean space with the usual orientation, namely $\omega = r^2 \sin\theta \, dr \wedge d\theta \wedge d\phi$.

   These are "physics" conventions: θ is the angle from the north pole (colatitude), and ϕ is the angle in the xy-plane (longitude).

   (a) Determine the Hodge dual operator $*$ on all forms (expressed in spherical coordinates) by computing its action on basis forms at each rank.

(b) Compute the dot and cross products of two arbitrary "vector fields" (really 1-forms) in spherical coordinates using the expressions

$$\alpha \cdot \beta = *(\alpha \wedge *\beta),$$
$$\alpha \times \beta = *(\alpha \wedge \beta).$$

You may express your results either with respect to an orthonormal basis or with respect to a "coordinate" (non-orthonormal) spherical basis; make sure you know which you're doing.

7. **Hodge Dual in Euclidean Space**

Consider Euclidean 4-space ($\mathbb{R}^4$) with coordinates (u, v, w, x), line element

$$ds^2 = du^2 + dv^2 + dw^2 + dx^2,$$

and orientation $\omega = du \wedge dv \wedge dw \wedge dx$.

(a) Determine the action of $*$ on all 2-forms.

(b) Find (nonzero) 2-forms S and A on $\mathbb{R}^4$ such that $*S = S$ and $*A = -A$. *(S and A are called **self-dual** and **anti-self-dual** 2-forms, respectively.)*

8. **Spacetime**

Consider the three-dimensional spacetime with coordinates t, θ, ϕ, line element

$$ds^2 = -dt^2 + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2\right),$$

and orientation $\omega = r^2 \sin\theta \, dt \wedge d\theta \wedge d\phi$, where r is a constant. Determine the action of $*$ on all differential forms on this spacetime.

≻ Chapter 15 ≺

Differentiation of Differential Forms

15.1 Gradient

We now have two algebraic maps on differential forms, namely,

$$\wedge : \bigwedge^p \times \bigwedge^q \longmapsto \bigwedge^{p+q}, \tag{15.1}$$

$$* : \bigwedge^p \longmapsto \bigwedge^{n-p}. \tag{15.2}$$

In this chapter, we introduce a third such map, involving differentiation.

Recall that $\bigwedge^0(M)$ is the set of functions on M, and that the differential df of any function f is a 1-form. Taking the differential, or "zapping a function with d," is therefore a map

$$d : \bigwedge^0 \longmapsto \bigwedge^1. \tag{15.3}$$

What is this map? We have

$$df = \frac{\partial f}{\partial x^i}\, dx^i, \tag{15.4}$$

a 1-form whose components are just the partial derivatives of f—just like the gradient. We therefore identify df with the gradient of f.

15.2 Exterior Differentiation

We would like to extend this notion of "gradient" to differential forms of higher rank. Any such form can be written in a coordinate basis as (a linear combination of terms like)

$$\alpha = f\, dx^I = f\, dx^{i_1} \wedge ... \wedge dx^{i_p}. \tag{15.5}$$

Just as taking the gradient of a function increases the rank by one, we would like to define $d\alpha$ to be a $p+1$-form, resulting in a map

$$d : \bigwedge^p \longmapsto \bigwedge^{p+1} . \tag{15.6}$$

The obvious place to add an extra d is again to take f to df, which suggests that we should set

$$d\alpha = df \wedge dx^I \tag{15.7}$$

and extend by linearity. Before summarizing the properties of this new operation, called *exterior differentiation*, we consider some examples.

15.3 DIVERGENCE AND CURL

Working in Euclidean $\mathbb{R}^3$, consider a 1-form

$$F = F_x\, dx + F_y\, dy + F_z\, dz. \tag{15.8}$$

What is dF? We have

$$\begin{aligned} d(F_x\, dx) &= dF_x \wedge dx \\ &= \left(\frac{\partial F_x}{\partial x}\, dx + \frac{\partial F_x}{\partial y}\, dy + \frac{\partial F_x}{\partial z}\, dz \right) \wedge dx \\ &= \frac{\partial F_x}{\partial z}\, dz \wedge dx - \frac{\partial F_x}{\partial y}\, dx \wedge dy. \end{aligned} \tag{15.9}$$

Adding up three such terms, we obtain

$$\begin{aligned} dF = \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) dy \wedge dz + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) dz \wedge dx \\ + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) dx \wedge dy. \end{aligned} \tag{15.10}$$

But these components look just like those of $\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}$, except that we have taken a 1-form to a 2-form! We can undo this using the Hodge dual. Thus, under the usual identification

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} \tag{15.11}$$

of vectors and 1-forms, we also have

$$dF = (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}) \cdot d\vec{\boldsymbol{A}} \tag{15.12}$$

or equivalently

$$*dF = (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}) \cdot d\vec{\boldsymbol{r}}. \tag{15.13}$$

Similarly, starting from a 2-form

$$G = F_x\, dy \wedge dz + F_y\, dz \wedge dx + F_z\, dx \wedge dy, \tag{15.14}$$

we ask what dG is, and compute

$$\begin{aligned} dG &= dF_x \wedge dy \wedge dz + dF_y \wedge dz \wedge dx + dF_z \wedge dx \wedge dy \\ &= \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \right) dx \wedge dy \wedge dz. \end{aligned} \tag{15.15}$$

This looks like $\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}$, except that it's a 3-form, namely

$$dG = (\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}})\, dV. \tag{15.16}$$

Furthermore, G is just $*F$. Putting this all together, we see that

$$*d{*F} = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}. \tag{15.17}$$

Finally, returning to the gradient, we can similarly write

$$df = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}}. \tag{15.18}$$

Thus, as suggested in Section 12.5, each of the vector calculus operators div, grad, and curl has an analog in the language of differential forms. Explicitly, if f is a function and F is the 1-form corresponding to the vector field $\vec{\boldsymbol{F}}$, then df is the 1-form corresponding to $\vec{\boldsymbol{\nabla}} f$, $*dF$ is the 1-form corresponding to $\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}$, and $*d{*F}$ is the function $\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}$. Furthermore, each of these operators is really just the exterior derivative d, acting as

$$d : \bigwedge^0 \longrightarrow \bigwedge^1 \longrightarrow \bigwedge^2 \longrightarrow \bigwedge^3. \tag{15.19}$$

15.4 Laplacian in Polar Coordinates

The *Laplacian* of a function f is defined by

$$\Delta f = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{\nabla}} f. \tag{15.20}$$

Rewriting the Laplacian in terms of differential forms, we obtain

$$\Delta f = *d{*df}. \tag{15.21}$$

As an example, consider polar coordinates. We have

$$\begin{aligned}
\Delta f &= {*d{*df}} \\
&= {*d*}\left(\frac{\partial f}{\partial r}\,dr + \frac{\partial f}{\partial \phi}\,d\phi\right) \\
&= {*d*}\left(\frac{\partial f}{\partial r}\,dr + \frac{1}{r}\frac{\partial f}{\partial \phi}\,r\,d\phi\right) \\
&= {*d}\left(\frac{\partial f}{\partial r}\,r\,d\phi - \frac{1}{r}\frac{\partial f}{\partial \phi}\,dr\right) \\
&= {*}\left(d\left(\frac{\partial f}{\partial r}\,r\right) \wedge d\phi - d\left(\frac{1}{r}\frac{\partial f}{\partial \phi}\right) \wedge dr\right) \\
&= {*}\left(\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial r}\,r\right)\,dr \wedge d\phi - \frac{\partial}{\partial \phi}\left(\frac{1}{r}\frac{\partial f}{\partial \phi}\right)\,d\phi \wedge dr\right) \\
&= {*}\left(\frac{1}{r}\left[\frac{\partial}{\partial r}\left(\frac{\partial f}{\partial r}\,r\right) + \frac{\partial}{\partial \phi}\left(\frac{1}{r}\frac{\partial f}{\partial \phi}\right)\right]\,dr \wedge r\,d\phi\right) \\
&= \frac{1}{r}\frac{\partial}{\partial r}\left(r\,\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \phi^2}.
\end{aligned} \tag{15.22}$$

A similar computation can be used to determine the Laplacian in any dimension and signature.

15.5 PROPERTIES OF EXTERIOR DIFFERENTIATION

We now investigate the properties of exterior differentiation.

Consider $d^2 f = d(df)$. We have

$$\begin{aligned}
d^2 f = d(df) &= d\left(\frac{\partial f}{\partial x^i}\,dx^i\right) \\
&= d\left(\frac{\partial f}{\partial x^i}\right) \wedge dx^i \\
&= \frac{\partial^2 f}{\partial x^j \partial x^i}\,dx^j \wedge dx^i = 0,
\end{aligned} \tag{15.23}$$

since mixed partial derivatives are independent of the order of differentiation, and the wedge product is antisymmetric. More formally, interchanging the dummy indices i and j changes the sign in the wedge product, but not the coefficient, thus demonstrating that $d^2 f$ is equal to its negative and must therefore be zero.

A similar argument works for differential forms of arbitrary rank. Suppose that

$$\alpha = f\,dx^I \tag{15.24}$$

so that

$$d\alpha = df \wedge dx^I. \tag{15.25}$$

Then

$$\begin{aligned} d^2\alpha = d(d\alpha) &= d\left(\frac{\partial f}{\partial x^i}\,dx^i \wedge dx^I\right) \\ &= d\left(\frac{\partial f}{\partial x^i}\right) \wedge dx^i \wedge dx^I \\ &= \frac{\partial^2 f}{\partial x^j \partial x^i}\,dx^j \wedge dx^i \wedge dx^I = 0. \end{aligned} \tag{15.26}$$

This result is often summarized by writing

$$d^2 = 0. \tag{15.27}$$

In three Euclidean dimensions, this rule reduces to well-known identities. We have

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{\nabla}} f \longleftrightarrow *d(df) = *ddf = 0, \tag{15.28}$$

$$\vec{\boldsymbol{\nabla}} \cdot (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}) \longleftrightarrow *d{*}(*dF) = *d{*}{*}dF = *ddF = 0. \tag{15.29}$$

Thus, the familiar identities $\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{\nabla}} f = 0$ and $\vec{\boldsymbol{\nabla}} \cdot (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}) = 0$ are both special cases of $d^2 = 0$.

Like all differential operators, d also satisfies a product rule. With α a p-form given by (15.24), suppose that

$$\beta = h\,dx^J \tag{15.30}$$

is a q-form. Then

$$\begin{aligned} d(\alpha \wedge \beta) &= d\left((f\,dx^I) \wedge (h\,dx^J)\right) \\ &= d\left(fh\,dx^I \wedge dx^J\right) \\ &= d(fh) \wedge dx^I \wedge dx^J \\ &= (h\,df + f\,dh) \wedge dx^J \wedge dx^J \\ &= h\,df \wedge dx^I \wedge dx^J + f\,dh \wedge dx^I \wedge dx^J \\ &= (df \wedge dx^I) \wedge h\,dx^J + (-1)^p f\,dx^I \wedge (dh \wedge dx^J) \\ &= d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta, \end{aligned} \tag{15.31}$$

where the factor of $(-1)^p$ arises from moving dh through the p-form dx^I.

15.6 PRODUCT RULES

What is the gradient of a product of functions? The ordinary product rule for differentials is

$$d(fg) = g\,df + f\,dg, \tag{15.32}$$

and under the correspondence

$$df = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}}, \tag{15.33}$$

we obtain the standard product rule

$$\vec{\boldsymbol{\nabla}}(fg) = g\,\vec{\boldsymbol{\nabla}} f + f\,\vec{\boldsymbol{\nabla}} g. \tag{15.34}$$

Similar rules can be derived for other operators. For instance, if $f \in \bigwedge^0$ and $\alpha \in \bigwedge^1$, then

$$*d(f\,\alpha) = *(df \wedge \alpha + f\,d\alpha) = *(df \wedge \alpha) + f\,*d\alpha, \tag{15.35}$$

and recalling that

$$*(F \wedge G) = (\vec{\boldsymbol{F}} \times \vec{\boldsymbol{G}}) \cdot d\vec{\boldsymbol{r}}, \tag{15.36}$$

$$*dF = (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}) \cdot d\vec{\boldsymbol{r}}, \tag{15.37}$$

we see that

$$\vec{\boldsymbol{\nabla}} \times (f\vec{\boldsymbol{G}}) = \vec{\boldsymbol{\nabla}} f \times \vec{\boldsymbol{G}} + f\,\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{G}}. \tag{15.38}$$

Similarly,

$$*d*(f\,\alpha) = *d(f\,*\alpha) = *(df \wedge *\alpha + f\,d*\alpha) = *(df \wedge *\alpha) + f\,*d*\alpha, \tag{15.39}$$

and recalling that

$$*(F \wedge *G) = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{G}}, \tag{15.40}$$

$$*d*F = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}, \tag{15.41}$$

we see that

$$\vec{\boldsymbol{\nabla}} \cdot (f\vec{\boldsymbol{G}}) = \vec{\boldsymbol{\nabla}} f \cdot \vec{\boldsymbol{G}} + f\,\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{G}}. \tag{15.42}$$

The product rules considered so far are standard results in vector calculus. Here's one that may be less familiar. Suppose $\alpha, \beta \in \bigwedge^1$. Then

$$\begin{aligned} *d*\left(*(\alpha \wedge \beta)\right) &= *d(\alpha \wedge \beta) \\ &= *(d\alpha \wedge \beta - \alpha \wedge d\beta) \\ &= *(\beta \wedge d\alpha - \alpha \wedge d\beta) \\ &= *(\beta \wedge **d\alpha - \alpha \wedge **d\beta) \\ &= *(\beta \wedge *(*d\alpha)) - *(\alpha \wedge *(*d\beta)). \end{aligned} \tag{15.43}$$

We have shown that

$$\vec{\boldsymbol{\nabla}} \cdot (\vec{\boldsymbol{F}} \times \vec{\boldsymbol{G}}) = \vec{\boldsymbol{G}} \cdot (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}}) - \vec{\boldsymbol{F}} \cdot (\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{G}}). \tag{15.44}$$

15.7 MAXWELL'S EQUATIONS I

Maxwell's equations are a system of coupled differential equations for the electric field $\vec{\boldsymbol{E}}$ and the magnetic field $\vec{\boldsymbol{B}}$. In traditional language, they take the form

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{E}} = 4\pi\rho, \tag{15.45}$$

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{B}} = 0, \tag{15.46}$$

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{E}} + \dot{\vec{\boldsymbol{B}}} = 0, \tag{15.47}$$

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{B}} - \dot{\vec{\boldsymbol{E}}} = 4\pi\vec{\boldsymbol{J}}, \tag{15.48}$$

where ρ is the charge density, $\vec{\boldsymbol{J}}$ is the current density, and dots denote time derivatives. Taking the divergence of the last equation, and using the first, leads to the *continuity equation*

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{J}} + \dot{\rho} = 0, \tag{15.49}$$

and making the Ansatz

$$\vec{\boldsymbol{B}} = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{A}}, \tag{15.50}$$

$$\vec{\boldsymbol{E}} = -\vec{\boldsymbol{\nabla}}\Phi - \dot{\vec{\boldsymbol{A}}} \tag{15.51}$$

automatically solves the middle two (source-free) equations.

15.8 MAXWELL'S EQUATIONS II

It is straightforward to translate Maxwell's equations into the language of differential forms in three (Euclidean) dimensions. We obtain

$$*d{*}E = 4\pi\rho, \tag{15.52}$$

$$*d{*}B = 0, \tag{15.53}$$

$$*dE + \dot{B} = 0, \tag{15.54}$$

$$*dB - \dot{E} = 4\pi J, \tag{15.55}$$

but it is customary to take the Hodge dual of both sides, resulting in

$$d{*}E = 4\pi\, {*}\rho, \tag{15.56}$$
$$d{*}B = 0, \tag{15.57}$$
$$dE + {*}\dot{B} = 0, \tag{15.58}$$
$$dB - {*}\dot{E} = 4\pi\, {*}J. \tag{15.59}$$

Differentiating the last equation, and using the first, brings the continuity equation to the form

$$*d{*}J + \dot{\rho} = 0, \tag{15.60}$$

and making the Ansatz

$$B = {*}dA, \tag{15.61}$$
$$E = -d\Phi - \dot{A} \tag{15.62}$$

automatically solves the middle two (source-free) equations.

15.9 Maxwell's Equations III

A remarkable simplification occurs if we rewrite Maxwell's equations using differential forms in four-dimensional Minkowski space, as we now show. We will assume that the orientation is given by

$$\omega = dx \wedge dy \wedge dz \wedge dt. \tag{15.63}$$

We first need to introduce some temporary notation, so that we can consider both three- and four-dimensional quantities in the same computation. We will therefore add a "bar" to all three-dimensional quantities. For 1-forms, such as $\bar{B}$ and $\bar{E}$, the bar indicates the absence of a dt-component. We also have the barred operators, $\bar{d}$ and $\bar{*}$, which ignore t.

It is straightforward to establish the following identities:

$$df = \bar{d}f + \dot{f}\, dt, \tag{15.64}$$
$$d\bar{E} = \bar{d}\bar{E} + dt \wedge \dot{\bar{E}}, \tag{15.65}$$
$$d\bar{*}\bar{E} = \bar{d}\bar{*}\bar{E} + dt \wedge \bar{*}\dot{\bar{E}}, \tag{15.66}$$
$$*\bar{E} = \bar{*}\bar{E} \wedge dt, \tag{15.67}$$
$$*(f\, dt) = \bar{*}f. \tag{15.68}$$

We can combine the 3 degrees of freedom of $\bar{E}$ with the 3 degrees of freedom of $\bar{B}$ into the 6 degrees of freedom in a (four-dimensional) 2-form, which we choose to do by defining

$$F = \bar{E} \wedge dt + \bar{*}\bar{B}, \tag{15.69}$$

and we note first of all that

$$*F = \bar{B} \wedge dt - \bar{*}\bar{E} \tag{15.70}$$

so that the Hodge dual effectively interchanges the electric and magnetic fields.

Direct computation now shows that

$$\begin{aligned} dF &= d\bar{E} \wedge dt + d\bar{*}\bar{B} \\ &= \bar{d}\bar{E} \wedge dt + \bar{d}\bar{*}\bar{B} + dt \wedge \bar{*}\dot{\bar{B}} \\ &= \bar{d}\bar{*}\bar{B} + (\bar{d}\bar{E} + \bar{*}\dot{\bar{B}}) \wedge dt. \end{aligned} \tag{15.71}$$

But the middle two of Maxwell's equations now imply that

$$dF = 0, \tag{15.72}$$

and the implication goes both ways, since the two summands in (15.71) are linearly independent. Similarly,

$$\begin{aligned} d{*}F &= d\bar{B} \wedge dt - d\bar{*}\bar{E} \\ &= \bar{d}\bar{B} \wedge dt - \bar{d}\bar{*}\bar{E} - dt \wedge \bar{*}\dot{\bar{E}} \\ &= -\bar{d}\bar{*}\bar{E} + (\bar{d}\bar{B} - \bar{*}\dot{\bar{E}}) \wedge dt, \end{aligned} \tag{15.73}$$

and the remaining two of Maxwell's equations bring this to the form

$$d{*}F = -4\pi\,\bar{*}\rho + 4\pi\,\bar{*}\bar{J} \wedge dt = -4\pi\,{*}(\rho\,dt) + 4\pi\,{*}\bar{J}. \tag{15.74}$$

We are thus led to define the *4-current density*

$$J = \bar{J} - \rho\,dt, \tag{15.75}$$

which combines the classical charge and current densities into a single object, and which brings the remaining two of Maxwell's equations to the form

$$d{*}F = 4\pi\,{*}J. \tag{15.76}$$

Similarly, combining the scalar and vector potentials Φ and $\bar{A}$ into a single *4-potential* via

$$A = \bar{A} - \Phi\, dt \tag{15.77}$$

results, perhaps not surprisingly, in

$$\begin{aligned} dA &= d\bar{A} - d(\Phi\, dt) \\ &= \bar{d}\bar{A} + dt \wedge \dot{A} - \bar{d}\Phi \wedge dt \\ &= F. \end{aligned} \tag{15.78}$$

In summary, Maxwell's equations in Minkowski space can be reduced to the two equations

$$F = dA, \tag{15.79}$$

$$d{*}F = 4\pi\, {*}J. \tag{15.80}$$

15.10 ORTHOGONAL COORDINATES

An *orthogonal* coordinate system is one in which the coordinate directions are mutually perpendicular. The standard examples are rectangular, polar, cylindrical, and spherical coordinates.

Working in Euclidean $\mathbb{R}^3$, suppose that (u, v, w) are orthogonal coordinates. Then an infinitesimal displacement $d\vec{\boldsymbol{r}}$ between nearby points can be expressed as

$$d\vec{\boldsymbol{r}} = h_u\, du\, \hat{\boldsymbol{u}} + h_v\, dv\, \hat{\boldsymbol{v}} + h_w\, dw\, \hat{\boldsymbol{w}} \tag{15.81}$$

for some functions h_u, h_v, h_w, where $\{\hat{\boldsymbol{u}}, \hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}\}$ is an orthonormal basis. Equivalently, the line element takes the form

$$ds^2 = d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = h_u^2\, du^2 + h_v^2\, dv^2 + h_w^2\, dw^2. \tag{15.82}$$

Since $d\vec{\boldsymbol{r}}$ expresses the infinitesimal displacement in mutually orthogonal directions, its components, namely $\{h_u\, du, h_v\, dv, h_w\, dw\}$, should be an orthonormal basis of 1-forms.[1]

We also have

$$d\vec{\boldsymbol{r}} = \frac{\partial \vec{\boldsymbol{r}}}{\partial u}\, du + \frac{\partial \vec{\boldsymbol{r}}}{\partial v}\, dv + \frac{\partial \vec{\boldsymbol{r}}}{\partial w}\, dw, \tag{15.83}$$

[1] This could be verified by expressing u, v, w in terms of x, y, z, substituting the result into ds^2, and comparing with the known result $ds^2 = dx^2 + dy^2 + dz^2$. The resulting constraints can be solved and inserted into du, dv, dw, expressed in terms of dx, dy, dz, and g can now be evaluated in this basis. This method is, however, equivalent to inverting a 3×3 matrix; we will learn an easier method in Section 16.1.

and comparison with (15.81) shows that

$$h_u\,\hat{\boldsymbol{u}} = \frac{\partial \vec{\boldsymbol{r}}}{\partial u}, \tag{15.84}$$

which in turn implies that

$$h_u = \left|\frac{\partial \vec{\boldsymbol{r}}}{\partial u}\right|, \tag{15.85}$$

$$\hat{\boldsymbol{u}} = \frac{1}{h_u}\frac{\partial \vec{\boldsymbol{r}}}{\partial u} \tag{15.86}$$

and similarly for v and w. Thus, if $\vec{\boldsymbol{r}}$ can be expressed in terms of u, v, w—the traditional parametrization—then the orthonormal bases (of both vectors and 1-forms) can be computed easily.

15.11 Div, Grad, Curl in Orthogonal Coordinates

What is the gradient in an orthogonal coordinate system? From

$$df = \frac{\partial h_u}{\partial u}\,du + \frac{\partial h_v}{\partial v}\,dv + \frac{\partial h_w}{\partial w}\,dw = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} \tag{15.87}$$

and

$$d\vec{\boldsymbol{r}} = h_u\,du\,\hat{\boldsymbol{u}} + h_v\,dv\,\hat{\boldsymbol{v}} + h_w\,dw\,\hat{\boldsymbol{w}}, \tag{15.88}$$

we see that

$$\vec{\boldsymbol{\nabla}} f = \frac{1}{h_u}\frac{\partial f}{\partial u}\hat{\boldsymbol{u}} + \frac{1}{h_v}\frac{\partial f}{\partial v}\hat{\boldsymbol{v}} + \frac{1}{h_w}\frac{\partial f}{\partial w}\hat{\boldsymbol{w}}. \tag{15.89}$$

What about divergence and curl? These formulas can of course be computed using d and $*$ in the usual way, but there is also a shortcut, which we present in the language of vector calculus.

Since

$$\vec{\boldsymbol{\nabla}} u = \frac{1}{h_u}\hat{\boldsymbol{u}}, \tag{15.90}$$

we see that

$$\vec{\boldsymbol{\nabla}} \times \left(\frac{1}{h_u}\hat{\boldsymbol{u}}\right) = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{\nabla}} u = \vec{\boldsymbol{0}}. \tag{15.91}$$

This means that $\{\frac{\hat{\boldsymbol{u}}}{h_u}, \frac{\hat{\boldsymbol{v}}}{h_v}, \frac{\hat{\boldsymbol{w}}}{h_w}\}$ is a *curl-free basis*, and computing the curl is easiest if we use this basis. We have

$$\begin{aligned}
\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} &= \vec{\boldsymbol{\nabla}} \times (F^u\,\hat{\boldsymbol{u}} + F^v\,\hat{\boldsymbol{v}} + F^w\,\hat{\boldsymbol{w}}) \\
&= \vec{\boldsymbol{\nabla}} \times \left((h_u F^u)\frac{\hat{\boldsymbol{u}}}{h_u} + (h_v F^v)\frac{\hat{\boldsymbol{v}}}{h_v} + (h_w F^w)\frac{\hat{\boldsymbol{w}}}{h_w}\right) \\
&= \vec{\boldsymbol{\nabla}}(h_u F^u) \times \frac{\hat{\boldsymbol{u}}}{h_u} + \vec{\boldsymbol{\nabla}}(h_v F^v) \times \frac{\hat{\boldsymbol{v}}}{h_v} + \vec{\boldsymbol{\nabla}}(h_w F^w) \times \frac{\hat{\boldsymbol{w}}}{h_w}.
\end{aligned} \tag{15.92}$$

Considering just the first term, we have

$$\begin{aligned}
\vec{\boldsymbol{\nabla}} \times (F^u\,\hat{\boldsymbol{u}}) &= \begin{vmatrix} \hat{\boldsymbol{u}} & \hat{\boldsymbol{v}} & \hat{\boldsymbol{w}} \\ \frac{1}{h_u}\frac{\partial}{\partial u}(h_u F^u) & \frac{1}{h_v}\frac{\partial}{\partial v}(h_u F^u) & \frac{1}{h_w}\frac{\partial}{\partial w}(h_u F^u) \\ \frac{1}{h_u} & 0 & 0 \end{vmatrix} \\
&= \frac{1}{h_u h_v h_w}\begin{vmatrix} h_u\,\hat{\boldsymbol{u}} & h_v\,\hat{\boldsymbol{v}} & h_w\,\hat{\boldsymbol{w}} \\ \frac{\partial}{\partial u}(h_u F^u) & \frac{\partial}{\partial v}(h_u F^u) & \frac{\partial}{\partial w}(h_u F^u) \\ 1 & 0 & 0 \end{vmatrix} \\
&= \frac{1}{h_u h_v h_w}\begin{vmatrix} h_u\,\hat{\boldsymbol{u}} & h_v\,\hat{\boldsymbol{v}} & h_w\,\hat{\boldsymbol{w}} \\ \frac{\partial}{\partial u} & \frac{\partial}{\partial v} & \frac{\partial}{\partial w} \\ h_u F^u & 0 & 0 \end{vmatrix}.
\end{aligned} \tag{15.93}$$

Putting back the missing terms, we obtain the following remarkable formula for curl in an arbitrary orthogonal coordinate system:

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} = \frac{1}{h_u h_v h_w}\begin{vmatrix} h_u\,\hat{\boldsymbol{u}} & h_v\,\hat{\boldsymbol{v}} & h_w\,\hat{\boldsymbol{w}} \\ \frac{\partial}{\partial u} & \frac{\partial}{\partial v} & \frac{\partial}{\partial w} \\ h_u F^u & h_v F^v & h_w F^w \end{vmatrix}. \tag{15.94}$$

We can proceed similarly for the divergence. Since

$$\vec{\boldsymbol{\nabla}} \times \left(v\,\vec{\boldsymbol{\nabla}} w\right) = \vec{\boldsymbol{\nabla}} v \times \vec{\boldsymbol{\nabla}} w = \frac{\hat{\boldsymbol{v}}}{h_v} \times \frac{\hat{\boldsymbol{w}}}{h_w} = \frac{\hat{\boldsymbol{u}}}{h_v h_w}, \tag{15.95}$$

we have

$$\vec{\boldsymbol{\nabla}} \cdot \left(\frac{\hat{\boldsymbol{u}}}{h_v h_w}\right) = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{\nabla}} \times \left(v\,\vec{\boldsymbol{\nabla}} w\right) = 0. \tag{15.96}$$

Thus, $\{\frac{\hat{\boldsymbol{u}}}{h_v h_w}, \frac{\hat{\boldsymbol{v}}}{h_w h_u}, \frac{\hat{\boldsymbol{w}}}{h_u h_v}\}$ is a *divergence-free basis*, which we can use to

simplify the computation of the divergence. We have

$$\begin{aligned}
\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}} &= \vec{\boldsymbol{\nabla}} \cdot (F^u\, \hat{\boldsymbol{u}} + F^v\, \hat{\boldsymbol{v}} + F^w\, \hat{\boldsymbol{w}}) \\
&= \vec{\boldsymbol{\nabla}} \cdot \left((h_v h_w F^u) \frac{\hat{\boldsymbol{u}}}{h_v h_w} + (h_w h_u F^v) \frac{\hat{\boldsymbol{v}}}{h_w h_u} + (h_u h_v F^w) \frac{\hat{\boldsymbol{w}}}{h_u h_v} \right) \\
&= \vec{\boldsymbol{\nabla}}(h_v h_w F^u) \cdot \frac{\hat{\boldsymbol{u}}}{h_v h_w} + \vec{\boldsymbol{\nabla}}(h_w h_u F^v) \cdot \frac{\hat{\boldsymbol{v}}}{h_w h_u} + \vec{\boldsymbol{\nabla}}(h_u h_v F^w) \cdot \frac{\hat{\boldsymbol{w}}}{h_u h_v} \\
&= \frac{1}{h_u h_v h_w} \left(\frac{\partial}{\partial u} (h_v h_w F^u) + \frac{\partial}{\partial v} (h_w h_u F^v) + \frac{\partial}{\partial w} (h_u h_v F^w) \right),
\end{aligned} \tag{15.97}$$

which we can rewrite in the form

$$\begin{aligned}
\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}} =& \frac{1}{h_u h_v h_w} \left(\hat{\boldsymbol{u}} \frac{\partial}{\partial u} + \hat{\boldsymbol{v}} \frac{\partial}{\partial v} + \hat{\boldsymbol{w}} \frac{\partial}{\partial w} \right) \\
&\cdot \ (h_v h_w F^u \hat{\boldsymbol{u}} + h_w h_u F^v \hat{\boldsymbol{v}} + \ h_u h_v F^w \hat{\boldsymbol{w}}) \,.
\end{aligned} \tag{15.98}$$

The Laplacian can now be computed as $\Delta f = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{\nabla}} f$.

It is instructive to verify formulas (15.94) and (15.98) using differential forms.

15.12 UNIQUENESS OF EXTERIOR DIFFERENTIATION

The properties of d in fact determine it uniquely.

Theorem: *There is a unique operator* $d : \bigwedge^p \longmapsto \bigwedge^{p+1}$ *satisfying the following properties:*

1. $d(a\,\alpha + \beta) = a\,d\alpha + d\beta$ *if* $a =$ *constant;*
2. $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p\, \alpha \wedge d\beta$ *if* α *is a p-form;*
3. $d^2\alpha = 0$*;*
4. $df = \frac{\partial f}{\partial x^i}\, dx^i$.

The first property is standard linearity, and the second is the product rule; all differential operators satisfy a linearity condition and a product rule. The last condition ensures that d does the right thing on functions, and could have been written

$$d(f) = df. \tag{15.99}$$

Existence is straightforward: Properties 1 and 4 are obvious, and the remaining two properties were established in Section 15.5.

To show uniqueness, we have to demonstrate that the defining equation for d follows from these properties. But the product rule with $p = 0$ takes the form

$$d(f\,\alpha) = df \wedge \alpha + f\,d\alpha \tag{15.100}$$

so that in particular

$$d(f\,dx^I) = df \wedge dx^I + f\,d\left(dx^I\right), \tag{15.101}$$

and Property 4 ensures that the first term is what we are looking for; it only remains to show that the last term is zero. But $d(dx^i) = 0$ by Property 3, after which $d(dx^i \wedge dx^j) = 0$ follows by Property 2. Continuing by induction, $d(dx^I) = 0$ for any I, and we are done.

15.13 PROBLEMS

1. **Product Rule I**

   Perform the indicated computations, including all reasonable simplifications.

   (a) Let f be a 0-form and suppose $df = \tau$. Determine $d(f\,\tau)$.

   (b) Let μ be a 1-form and suppose $d\mu = \nu$. Determine $d(\mu \wedge \nu)$.

2. **Identities**

   Derive the following identities.

   (a) Let $\vec{\boldsymbol{u}}$, $\vec{\boldsymbol{v}}$, $\vec{\boldsymbol{w}}$ be vectors in (Euclidean) $\mathbb{R}^3$. Show that

   $$(\vec{\boldsymbol{u}} \times \vec{\boldsymbol{v}}) \cdot \vec{\boldsymbol{w}} = \vec{\boldsymbol{u}} \cdot (\vec{\boldsymbol{v}} \times \vec{\boldsymbol{w}}).$$

   (b) Let h be a (smooth) function in (Euclidean) $\mathbb{R}^3$. Show that

   $$\vec{\boldsymbol{\nabla}} \cdot (h\vec{\boldsymbol{\nabla}}h) = h\,\triangle h + |\vec{\boldsymbol{\nabla}}h|^2.$$

3. **Product Rule II**

   Let h be a smooth function on $\mathbb{R}^3$ (Euclidean 3-space).

   (a) Use the product rule to expand $*d*(h\,dh)$.

   (b) Interpret the result as a product rule involving $\vec{\boldsymbol{\nabla}}$.

   (c) Does your result hold in other signatures and dimensions?

4. **Derivatives**

   Perform the indicated computations, including all reasonable simplifications.

   (a) Let α be a 1-form and suppose $d\alpha = \beta$. Find $d(\alpha \wedge \beta)$.

   (b) Let an orientation ω be given. What is $d\omega$? Why?

5. **Polar Coordinates**

   Consider ordinary polar coordinates, with line element

$$ds^2 = dr^2 + r^2 d\phi^2.$$

   Let $\rho = \ln r$, and consider the new coordinates (ρ, ϕ).

   (a) Express the line element in these coordinates.

   (b) Find an orthonormal basis of 1-forms adapted to these coordinates.

   (c) Express the standard orientation (ω) in terms of your basis.

   (d) Determine the Hodge dual $(*)$ of each 1-form in your basis.

   (e) Find the Laplacian $\triangle f = {*d{*df}}$ of a function f in these coordinates.

6. **Laplacian**

   Consider Euclidean 2-space $(\mathbb{R}^2)$ with coordinates (u, v) and line element

$$ds^2 = h^2\,(du^2 + dv^2)$$

   for some function $h = h(u, v)$.

   (a) Choose an orientation, that is, write down a possible choice of ω.

   (b) Determine the Laplacian ${*d{*df}}$ of an unknown function $f = f(u, v)$.

7. **Orthogonal Coordinates**

Choose any orthogonal coordinate system in three-dimensional Euclidean space $\mathbb{R}^3$ *other than* rectangular, cylindrical, or spherical coordinates. Working in an orthonormal basis, compute the gradient and Laplacian of an arbitrary function, and the curl and divergence of an arbitrary "vector field" (again, really a 1-form), using the expressions:

$$\begin{aligned} \nabla f &= df, \\ \nabla \times \alpha &= {*d\alpha}, \\ \nabla \cdot \alpha &= {*d{*\alpha}}, \\ \triangle f = \nabla \cdot \nabla f &= {*d{*df}}. \end{aligned}$$

Some other orthogonal coordinate systems in $\mathbb{R}^3$ *are given below, in each case expressed in terms of rectangular coordinates* (x, y, z) *(and with* a = *constant).*

Parabolic Cylindrical Coordinates:

$$x = \frac{1}{2}(u^2 - v^2), \quad y = uv, \quad z = z.$$

Paraboloidal Coordinates:

$$x = uv\cos\phi, \quad y = uv\sin\phi, \quad z = \frac{1}{2}(u^2 - v^2).$$

Elliptic Cylindrical Coordinates:

$$x = a\cosh u\cos v, \quad y = a\sinh u\sin v, \quad z = z.$$

Prolate Spheroidal Coordinates:

$$x = a\sinh\xi\sin\eta\cos\phi, \quad y = a\sinh\xi\sin\eta\sin\phi, \quad z = a\cosh\xi\cos\eta.$$

Oblate Spheroidal Coordinates:

$$x = a\cosh\xi\cos\eta\cos\phi, \quad y = a\cosh\xi\cos\eta\sin\phi, \quad z = a\sinh\xi\sin\eta.$$

Bipolar Coordinates:

$$x = \frac{a\sinh v}{\cosh v - \cos u}, \quad y = \frac{a\sin u}{\cosh v - \cos u}, \quad z = z.$$

Toroidal Coordinates:

$$x = \frac{a\sinh v\cos\phi}{\cosh v - \cos u}, \quad y = \frac{a\sinh v\sin\phi}{\cosh v - \cos u}, \quad z = \frac{a\sin u}{\cosh v - \cos u}.$$

≻ CHAPTER 16 ≺

INTEGRATION OF DIFFERENTIAL FORMS

16.1 VECTORS AND DIFFERENTIAL FORMS

We have repeatedly made use of the obvious correspondence between vector fields and 1-forms, under which $\hat{\boldsymbol{x}}$ becomes dx, etc. It is time to make this correspondence more formal. Throughout the following discussion, it is useful to imagine that we are working in ordinary three-dimensional Euclidean space, with the usual dot product. However, the argument is the same in any dimension and signature.

Suppose $\vec{\boldsymbol{F}}$ is a vector field. Then the corresponding 1-form is given by

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}. \tag{16.1}$$

This map is clearly an isomorphism of vector spaces (really modules), and therefore has an inverse, which we write in the form

$$g(F, d\vec{\boldsymbol{r}}) = \vec{\boldsymbol{F}} \tag{16.2}$$

and which can be taken as the definition of g. Assuming linearity and picking out the components of the vectors on both sides of the equality, (16.2) does indeed reduce to the inner product g on 1-forms that we have been using all along. Furthermore, the construction (16.2) clearly preserves the inner product, that is,

$$g(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}, \vec{\boldsymbol{G}} \cdot d\vec{\boldsymbol{r}}) = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{G}}, \tag{16.3}$$

which could also have been taken as the definition of g acting on 1-forms.

All three of these equations, (16.1), (16.2), and (16.3), make clear the fundamental role played by $d\vec{\boldsymbol{r}}$. Among other things, if $d\vec{\boldsymbol{r}}$ is expressed in terms of an orthonormal basis of vectors, $\{\hat{\boldsymbol{e}}_i\}$, that is, if

$$d\vec{\boldsymbol{r}} = \sigma^i \, \hat{\boldsymbol{e}}_i \tag{16.4}$$

for some 1-forms σ^i, then from (16.3) we have $\sigma^i = \pm\hat{\boldsymbol{e}}_i \cdot d\vec{\boldsymbol{r}}$ and

$$g(\sigma^i, \sigma^j) = g(\hat{\boldsymbol{e}}_i \cdot d\vec{\boldsymbol{r}}, \hat{\boldsymbol{e}}_j \cdot d\vec{\boldsymbol{r}},) = \hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j = \pm\delta_{ij} \tag{16.5}$$

so that $\{\sigma^i\}$ is an orthonormal basis of 1-forms, a fact we have already been using.

16.2 LINE AND SURFACE INTEGRALS

We began our study of differential forms by claiming that differential forms are integrands. Which ones?

Using the results of Section 16.1, line integrals are easy. We have

$$\int_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} = \int_C F, \tag{16.6}$$

where of course

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}} \tag{16.7}$$

is the 1-form corresponding to $\vec{\boldsymbol{F}}$.

What about surface integrals? Consider first some simple examples. What is the flux of $\vec{\boldsymbol{F}}$ upward through the xy-plane? We have

$$\int_S \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = \int_S \vec{\boldsymbol{F}} \cdot \hat{\boldsymbol{z}}\, dx\, dy = \int_S F_z\, dx\, dy, \tag{16.8}$$

where F_z is the z-component of $\vec{\boldsymbol{F}}$. The integrand on the right-hand side can be reinterpreted as $F_z\, dx \wedge dy$, which is one of the components of

$$*F = F_x\, dy \wedge dz + F_y\, dz \wedge dx + F_z\, dx \wedge dy. \tag{16.9}$$

But since $z = 0$ on S, $dz = 0$ there, and we have

$$\int_S \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = \int_S *F. \tag{16.10}$$

A similar argument in spherical coordinates for the flux of $\vec{\boldsymbol{F}}$ out of a sphere centered at the origin results in

$$\int_S \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = \int_S F_r\, r^2 \sin\theta\, d\theta\, d\phi, \tag{16.11}$$

and again the integrand is just $*F$ *restricted to the sphere*, since r being a constant implies that $dr = 0$. We are thus led to conjecture that

$$*F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}. \tag{16.12}$$

One more example should suffice.

Consider the graph of a function of the form $z = f(x, y)$. What is the surface element on such a surface? As shown in Figure 16.1, we chop the

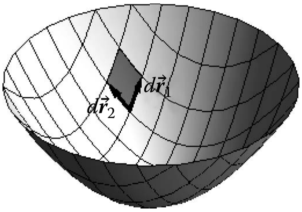


FIGURE 16.1. Chopping up a surface in rectangular coordinates.

surface along curves with $y =$ constant and $x =$ constant, and consider $d\vec{\boldsymbol{r}}$ along such curves, obtaining

$$d\vec{\boldsymbol{r}}_1 = \left(\hat{\boldsymbol{x}} + \frac{\partial z}{\partial x}\,\hat{\boldsymbol{z}}\right) dx, \tag{16.13}$$

$$d\vec{\boldsymbol{r}}_2 = \left(\hat{\boldsymbol{y}} + \frac{\partial z}{\partial y}\,\hat{\boldsymbol{z}}\right) dy \tag{16.14}$$

so that

$$d\vec{\boldsymbol{A}} = d\vec{\boldsymbol{r}}_1 \times d\vec{\boldsymbol{r}}_2 = \left(-\frac{\partial z}{\partial x}\,\hat{\boldsymbol{x}} - \frac{\partial z}{\partial y}\,\hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}\right) dx\,dy \tag{16.15}$$

and

$$\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = \left(-F_x\frac{\partial z}{\partial x} - F_y\frac{\partial z}{\partial y} + F_z\right) dx\,dy. \tag{16.16}$$

But

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy \tag{16.17}$$

so that

$$\begin{aligned} {*F} &= F_x\,dy\wedge dz + F_y\,dz\wedge dx + F_z\,dx\wedge dy \\ &= \left(-F_x\frac{\partial z}{\partial x} - F_y\frac{\partial z}{\partial y} + F_z\right) dx\wedge dy \end{aligned} \tag{16.18}$$

when restricted to the surface.

The above examples were all three-dimensional, but the argument is the same in n dimensions (and arbitrary signature, although the orientation must be checked). Line integrals are just as you'd expect, but surface integrals now correspond to the flux through an $(n-1)$-dimensional surface—also what you would expect, since $*F$ is an $(n-1)$-form.

16.3 Integrands Revisited

We have seen that (in $\mathbb{R}^3$)

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}, \tag{16.19}$$

$$*F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} \tag{16.20}$$

when we integrate both sides. This restriction leads us to a slightly different interpretation of integrands than you may be used to. Integrands by themselves can be thought of as being defined everywhere; it is only after you decide where to integrate them that they are restricted to the domain of integration. If two integrands are equal after being integrated over *any* domain, we say that the integrands themselves are equal. It is in this sense that equality holds in expressions such as (16.19) and (16.20).

So much for line and surface integrals. What about volume integrals? The volume element dV in n dimensions is just the orientation $\omega = *1$, so we have

$$f\,dV = f\,\omega = f *1 = *f \tag{16.21}$$

or simply

$$*f = f\,dV. \tag{16.22}$$

Finally, since we now know how to integrate F, $*F$, and $*f$, it seems natural to ask whether we can integrate f. Be careful! This is f the 0-form, not $f\,dx$ the 1-form! A 0-form must be integrated over a zero-dimensional "surface," that is, over one or more points. So we *define*

$$\int_p f = f(p). \tag{16.23}$$

16.4 Stokes' Theorem

There are three "big theorems" in vector calculus, namely the *Divergence Theorem*,

$$\int_R \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}\,dV = \oint_{\partial R} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}; \tag{16.24}$$

Stokes' Theorem,

$$\int_S \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = \oint_{\partial S} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}; \tag{16.25}$$

and

$$\int_A^B \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} = f\Big|_A^B \tag{16.26}$$

where ∂ stands for "the boundary of" and $\oint$ denotes integration over a closed domain. This latter theorem is sometimes referred to as the fundamental theorem for the gradient, just as the first two can be thought of as fundamental theorems for the divergence and curl, respectively. All three of these theorems are based on the fundamental theorem of calculus, which says

$$\int_a^b \frac{df}{dx}\,dx = f(b) - f(a). \tag{16.27}$$

We rewrite each of these theorems in the language of differential forms. Beginning with (16.26), we have

$$\int_C df = \int_{\partial C} f, \tag{16.28}$$

where C is any path from A to B and where we have used the integral notation introduced in Section 16.3 for function evaluation. Turning to (16.25), we have

$$\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = *\big(*dF\big) = dF, \tag{16.29}$$

so Stokes' Theorem becomes

$$\int_S dF = \int_{\partial S} F. \tag{16.30}$$

Finally, turning to (16.24), we have

$$\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}\,dV = *\big(*d{*}F\big) = d{*}F, \tag{16.31}$$

and the Divergence Theorem becomes

$$\int_R d{*}F = \int_{\partial R} {*}F. \tag{16.32}$$

Comparing these three results, we can conjecture that for *any* differential form α, in *any* dimension and signature we must have

$$\int_R d\alpha = \int_{\partial R} \alpha, \tag{16.33}$$

which turns out to be correct. The proof of this result, also called *Stokes' Theorem*, is beyond the scope of this book, although the only real subtlety is keeping track of the relative orientations of R and its boundary ∂R.

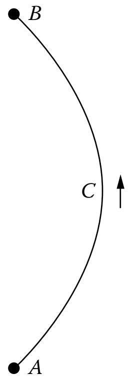


FIGURE 16.2. A curved path C from A to B.

16.5 CALCULUS THEOREMS

We restate the results in Section 16.4, reversing the order. *Stokes' Theorem* for differential forms says that

$$\int_R d\alpha = \int_{\partial R} \alpha \tag{16.34}$$

for any p-form α and any $(p+1)$-dimensional region R. All of the standard theorems in calculus relating integrals over regions and their boundaries are special cases of Stokes' Theorem.

FUNDAMENTAL THEOREM OF CALCULUS

The Fundamental Theorem of Calculus is about integrating the derivative of a function along a curve, namely,

$$\int_C df = \int_{\partial C} f = f\Big|_A^B, \tag{16.35}$$

where the curve C starts at point A and ends at point B, as shown in Figure 16.2. This theorem includes both the single-variable statement that

$$\int_a^b \frac{df}{dx}\,dx = f(b) - f(a), \tag{16.36}$$

in which case the curve lies along the x-axis, and the fundamental theorem for the gradient, namely

$$\int_C \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} = f\Big|_A^B, \tag{16.37}$$

where C is any smooth curve.

GREEN'S THEOREM

Green's Theorem is about integrating vector fields in the plane. If $F = F_x\,dx + F_y\,dy$, then

$$dF = \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) dx \wedge dy \tag{16.38}$$

so that

$$\int_S dF = \int_{\partial S} F \Longrightarrow \int_S \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) dA = \oint_{\partial S} F_x\,dx + F_y\,dy, \tag{16.39}$$

which is the special case of Stokes' Theorem restricted to the plane.

STOKES' THEOREM

Stokes' Theorem is about integrating the curl of a vector field. If $F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$, then

$$dF = *(*dF) = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} \tag{16.40}$$

so that

$$\int_S dF = \int_{\partial S} F \Longrightarrow \int_S \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}} = \oint_{\partial S} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}. \tag{16.41}$$

The surface S and its boundary ∂S are shown in Figure 16.3.

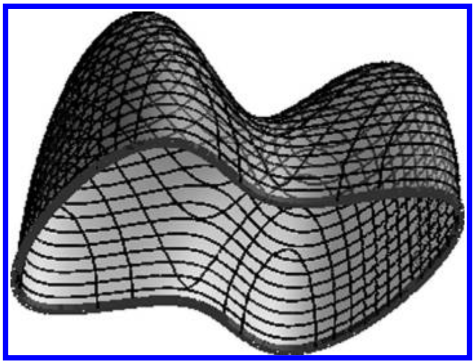


FIGURE 16.3. The geometry of Stokes' Theorem: The surface S is a three-dimensional "butterfly net"; the curve ∂S is its rim.

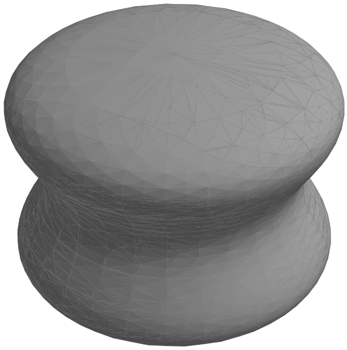


FIGURE 16.4. The geometry of the Divergence Theorem: The surface ∂R is the dumbbell-shaped surface shown; the region R is its interior.

DIVERGENCE THEOREM

The Divergence Theorem is about integrating the divergence of a vector field. If $F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$, then

$$*F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}, \tag{16.42}$$

$$d{*}F = *(*d{*}F) = \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}\, dV \tag{16.43}$$

so that

$$\int_R d{*}F = \int_{\partial R} *F \Longrightarrow \int_R \vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}}\, dV = \oint_{\partial R} \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}. \tag{16.44}$$

The region R and its boundary ∂R are shown in Figure 16.4.

16.6 INTEGRATION BY PARTS

Recall the product rule

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta \tag{16.45}$$

for a p-form α and q-form β. Rearranging and integrating this expression, we obtain

$$\begin{aligned} \int_R d\alpha \wedge \beta &= \int_R d(\alpha \wedge \beta) - (-1)^p \int_R \alpha \wedge d\beta \\ &= \int_{\partial R} \alpha \wedge \beta - (-1)^p \int_R \alpha \wedge d\beta. \end{aligned} \tag{16.46}$$

A special case is of course the usual formula for integration by parts, namely

$$\int_C f\,dg = \int_C (dg\, f) = \int_{\partial C} g\, f - \int_C g\, df, \tag{16.47}$$

and a more interesting example is

$$\int_D f\, d\alpha = \int_{\partial D} f\,\alpha - \int_D df \wedge \alpha \tag{16.48}$$

for any function f and p-form α.

16.7 COROLLARIES OF STOKES' THEOREM

If M is compact (and open), it has no boundary, so $\partial M = \emptyset$. The boundary of a disk is a circle, but a circle has no boundary. Similarly, the boundary of a ball is a sphere, but a sphere has no boundary. Circles and spheres are therefore examples of compact surfaces with no boundary.

If M has no boundary, then

$$\int_M d\alpha = \int_{\partial M} \alpha = 0 \tag{16.49}$$

for *any* $(n-1)$-form α. The contrapositive of this statement says that if an n-form β satisfies

$$\int_M \beta \neq 0, \tag{16.50}$$

then $\beta \neq d\alpha$ for any $(n-1)$-form α; as discussed in Section 20.4, we say that β is not *exact*. For example, the orientation itself satisfies

$$\int_M \omega > 0, \tag{16.51}$$

which shows that ω is not exact.

Similarly, suppose that an $(n-1)$-form α is such that

$$d\alpha = h\,\omega \tag{16.52}$$

for some function h. Since the integral of the left-hand side is zero, and the integral of ω is nonzero, either h is identically zero or it must change sign; in either case, $h = 0$ somewhere.

As an example, *all* smooth functions f on a circle must satisfy $df = 0$ somewhere. And any smooth vector field $\vec{\boldsymbol{F}}$ on a sphere must have vanishing curl somewhere, since $F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$ implies that $dF = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{A}}$.

16.8 PROBLEMS

1. **The 3-Sphere**

   The line element on the 3-sphere $\mathbb{S}^3$ of radius r is given by

   $$ds^2 = r^2 d\psi^2 + r^2 \sin^2\psi \left(d\theta^2 + \sin^2\theta \, d\phi^2\right),$$

   where $\psi, \theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Compute the volume of $\mathbb{S}^3$.
   This is the three-dimensional volume of the ***surface*** *of the 3-sphere, not the four-dimensional "hypervolume" of the solid ball inside it in* $\mathbb{R}^4$.

2. **Spheres**

   The line element in spherical coordinates in $\mathbb{R}^4$ is given by

   $$ds^2 = dr^2 + r^2 d\psi^2 + r^2 \sin^2\psi \left(d\theta^2 + \sin^2\theta \, d\phi^2\right),$$

   where $\psi, \theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Find the flux of the radial vector field $\vec{\boldsymbol{F}} = \frac{\hat{\boldsymbol{r}}}{r^3}$ in $\mathbb{R}^4$ outward through the 3-sphere $\mathbb{S}^3$ given by $r = a$.

3. **Laplacians**

   The Laplacian of a function f on the 2-sphere $\mathbb{S}^2$ is given by

   $$\triangle f = \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \, \frac{\partial f}{\partial\theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 f}{\partial^2 \phi}.$$

   (a) Guess a (well-behaved, nonzero) solution f_1 of $\triangle f = 0$ on $\mathbb{S}^2$. Then try to guess a fundamentally different solution f_2. Make a conjecture about the solutions of $\triangle f = 0$ on $\mathbb{S}^2$.

   (b) Let h be an unknown (well-behaved) function on $\mathbb{S}^2$. Integrate $h\, d(*dh)$ by parts over all of $\mathbb{S}^2$. Evaluate any terms you can.
   Expand $d(h *dh)$.

   (c) Assuming now that $\triangle h = 0$ on $\mathbb{S}^2$, what can you conclude about dh? About h?
   The (Euclidean) signature matters.

   (d) Was your conjecture correct?

4. **Integration**

   The line element for the *flat torus* $\mathbb{T}$ is given by

   $$ds^2 = a^2 d\theta^2 + b^2 d\phi^2,$$

   where a and b are constants, and $\theta, \phi \in [0, 2\pi]$. Let f be an unknown function on $\mathbb{T}$ that is well behaved, that is, differentiable (and hence continuous).

(a) Integrate $f\,d(*df)$ by parts over all of $\mathbb{T}$.
Start with $d(f * df)$.

(b) Assume now that $\triangle f = 0$ on $\mathbb{T}$, and make any further obvious simplifications.

(c) What can you conclude about df? About f? About solutions to $\triangle f = 0$ on the torus?
The (Euclidean) signature matters.

≻ CHAPTER 17 ≺

CONNECTIONS

17.1 POLAR COORDINATES II

We begin this chapter with a motivating example. How do $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\phi}}$ vary as you move from point to point? There are several ways to answer this question.

A geometric answer can be constructed by drawing $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\phi}}$ at two nearby points, then comparing them. If the points are separated radially, it is clear that these basis vectors do not change as you move from point to point. Thus, along such a path,

$$d\hat{\boldsymbol{r}} = 0 = d\hat{\boldsymbol{\phi}}. \tag{17.1}$$

What if you move through an angle $d\phi$ along a circle? It is now readily apparent that each basis vector is rotated through the same angle. Both of these possibilities are shown in Figure 17.1. Furthermore, the difference between two vectors of the same magnitude is a secant line on the circle connecting their tips, as shown in Figure 17.2 for $\hat{\boldsymbol{r}}$; a similar construction holds for $\hat{\boldsymbol{\phi}}$. As the two vectors approach each other, this secant line becomes a tangent line and is therefore perpendicular to the original vector.

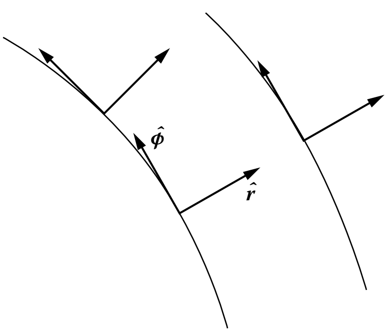


FIGURE 17.1. The polar basis vectors $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$ at three nearby points.

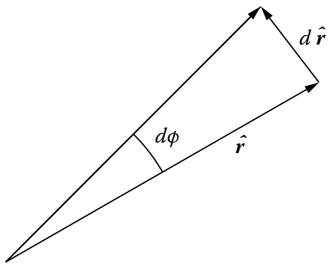


FIGURE 17.2. The change in $\hat{\boldsymbol{r}}$ in the ϕ direction.

(An algebraic proof of this result is given in the discussion of (17.9) below.) The magnitude of the difference between the two vectors is, in the limit, just the length of the arc between them, which, for unit vectors, is given by the angle. Thus, along such a path,

$$d\hat{\boldsymbol{r}} = d\phi\,\hat{\boldsymbol{\phi}}, \tag{17.2}$$

$$d\hat{\boldsymbol{\phi}} = -d\phi\,\hat{\boldsymbol{r}}, \tag{17.3}$$

and these expressions in fact hold along any path, since radial motion does not contribute.

This answer can also be obtained by converting to rectangular basis vectors, which are constant; it is easy to differentiate vector fields in this basis! We have

$$\hat{\boldsymbol{r}} = r\cos\phi\,\hat{\boldsymbol{x}} + r\sin\phi\,\hat{\boldsymbol{y}}, \tag{17.4}$$

$$\hat{\boldsymbol{\phi}} = -r\sin\phi\,\hat{\boldsymbol{x}} + r\cos\phi\,\hat{\boldsymbol{y}}, \tag{17.5}$$

from which (17.2) and (17.3) easily follow by differentiation.

A somewhat more sophisticated argument begins by differentiating

$$\vec{\boldsymbol{r}} = r\,\hat{\boldsymbol{r}} \tag{17.6}$$

to obtain

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + r\,d\hat{\boldsymbol{r}} \tag{17.7}$$

and then comparing this result with the known (geometrically derived) formula

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}}, \tag{17.8}$$

which verifies (17.2) immediately. We can determine $d\hat{\boldsymbol{\phi}}$ by noticing that

$$0 = d(\hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{\phi}}) = 2\,\hat{\boldsymbol{\phi}} \cdot d\hat{\boldsymbol{\phi}} \tag{17.9}$$

so that, first of all, $d\hat{\boldsymbol{\phi}}$ has no $\hat{\boldsymbol{\phi}}$ component. More generally, this argument shows that any derivative of a vector of constant length must be perpendicular to the original vector. Furthermore, we have

$$0 = d(\hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{r}}) = d\hat{\boldsymbol{\phi}} \cdot \hat{\boldsymbol{r}} + \hat{\boldsymbol{\phi}} \cdot d\hat{\boldsymbol{r}}, \tag{17.10}$$

which tells us that the $\hat{\boldsymbol{r}}$ component of $d\hat{\boldsymbol{\phi}}$ must be $-d\phi$. Putting these two facts together yields (17.3), as expected.

17.2 DIFFERENTIAL FORMS THAT ARE ALSO VECTOR FIELDS

The title says it all: We now consider differential forms that are also vector fields, which are called *vector-valued differential forms*. The standard example is $d\vec{\boldsymbol{r}}$ itself, which is both a 1-form and a vector field. More generally, a *vector-valued p-form* can be written as $\alpha^i \hat{\boldsymbol{e}}_i$, where each α^i is a p-form and where $\{\hat{\boldsymbol{e}}_i\}$ is a vector basis (here chosen to be orthonormal).

17.3 EXTERIOR DERIVATIVES OF VECTOR FIELDS

A vector field $\vec{\boldsymbol{v}}$ can be thought of as a vector-valued 0-form. As such, we should be able to take its exterior derivative and obtain a vector-valued 1-form $d\vec{\boldsymbol{v}}$.

The simplest examples occur when $\vec{\boldsymbol{v}}$ can be expanded in terms of a rectangular basis, e.g.,

$$\vec{\boldsymbol{F}} = F^x\,\hat{\boldsymbol{x}} + F^y\,\hat{\boldsymbol{y}} \tag{17.11}$$

in two dimensions. Surely, since $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are constant vector fields, we should have

$$d\hat{\boldsymbol{x}} = 0 = d\hat{\boldsymbol{y}} \tag{17.12}$$

so that

$$d\vec{\boldsymbol{F}} = dF^x\,\hat{\boldsymbol{x}} + dF^y\,\hat{\boldsymbol{y}}. \tag{17.13}$$

In an arbitrary basis, we would have

$$\vec{\boldsymbol{F}} = F^i\,\hat{\boldsymbol{e}}_i \tag{17.14}$$

and now

$$d\vec{\boldsymbol{F}} = dF^i\,\hat{\boldsymbol{e}}_i + F^i\,d\hat{\boldsymbol{e}}_i. \tag{17.15}$$

Thus, to determine the exterior derivative of vector fields, we need to know how to take the exterior derivative of our basis vectors.

17.4 PROPERTIES OF DIFFERENTIATION

So what properties should exterior differentiation of vector fields satisfy? Like all derivative operators, we must clearly demand linearity,

$$d(a\,\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}) = a\,d\vec{\boldsymbol{v}} + d\vec{\boldsymbol{w}}, \tag{17.16}$$

where a is a constant, and a product rule,

$$d(\alpha\,\vec{\boldsymbol{v}}) = d\alpha\,\vec{\boldsymbol{v}} + (-1)^p \alpha \wedge d\vec{\boldsymbol{v}}, \tag{17.17}$$

where α is a p-form. An alternative version of the product rule is obtained by reversing the order of the factors, namely

$$d(\vec{\boldsymbol{v}}\,\alpha) = d\vec{\boldsymbol{v}} \wedge \alpha + \vec{\boldsymbol{v}}\,d\alpha, \tag{17.18}$$

which has no minus signs, since $\vec{\boldsymbol{v}}$ is a (vector-valued) 0-form.

Do these properties suffice to determine the action of d on vector-valued differential forms? No! We need more information about d acting on a vector basis.

17.5 CONNECTIONS

Let's turn this problem on its head. We don't yet know how d acts on our basis $\{\hat{\boldsymbol{e}}_i\}$, but let's give this action a name. We must have

$$d\hat{\boldsymbol{e}}_j = \omega^i{}_j\,\hat{\boldsymbol{e}}_i \tag{17.19}$$

for some 1-forms $\omega^i{}_j$, which are called *connection 1-forms*. In fact, *any* choice of these 1-forms determines an exterior derivative operator satisfying the two conditions (17.16) and (17.17). As we will see in Section 17.6, further conditions must be imposed in order to determine a preferred connection.

We establish some notation. First of all, let

$$g_{ij} = \hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j. \tag{17.20}$$

In an orthonormal basis, the matrix (g_{ij}) will of course be diagonal, with diagonal entries ± 1. The g_{ij} are the components of an object called the *metric tensor*, which is closely related to the line element, since

$$ds^2 = d\vec{\boldsymbol{r}} \cdot d\vec{\boldsymbol{r}} = \big(\sigma^i\,\hat{\boldsymbol{e}}_i\big) \cdot \big(\sigma^j\,\hat{\boldsymbol{e}}_j\big) = g_{ij}\,\sigma^i\,\sigma^j. \tag{17.21}$$

We then set

$$\omega_{ij} = \hat{\boldsymbol{e}}_i \cdot d\hat{\boldsymbol{e}}_j \tag{17.22}$$

so that we also have

$$\omega_{ij} = g_{ik}\, \omega^k{}_j. \tag{17.23}$$

In Euclidean signature, (g_{ij}) is the identity matrix and there is no difference between "up" and "down" indices, but in other signatures there will be sign differences, so it is important to keep track.

17.6 THE LEVI-CIVITA CONNECTION

We now impose two additional conditions on the connection, then argue that these are enough to determine the connection uniquely.

First, we require that

$$d(\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}}) = d\vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{w}} + \vec{\boldsymbol{v}} \cdot d\vec{\boldsymbol{w}} \tag{17.24}$$

so that differentiation respects the dot product. Applying this condition to our orthonormal bases, we have

$$0 = d(\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j) = d\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j + \hat{\boldsymbol{e}}_i \cdot d\hat{\boldsymbol{e}}_j = \omega_{ji} + \omega_{ij}. \tag{17.25}$$

A connection that satisfies (17.25) is called *metric compatible*.

Next, we consider d^2 on vectors. We would like to demand that this be zero, but that turns out to be too strong. Consider, for example, $d\vec{\boldsymbol{r}}$. In the standard Euclidean spaces of vector calculus, it is clear that

$$d\vec{\boldsymbol{r}} = d(\vec{\boldsymbol{r}}), \tag{17.26}$$

where $\vec{\boldsymbol{r}}$ is the position vector, and it is then reasonable to expect that $d^2\vec{\boldsymbol{r}} = 0$. But on the surface of a sphere, for example, $d\vec{\boldsymbol{r}}$ makes perfect sense, but $\vec{\boldsymbol{r}}$ doesn't exist! In this context, $d\vec{\boldsymbol{r}}$ is not "d" of anything, so it is not obvious what $d^2\vec{\boldsymbol{r}}$ should be.

We will nonetheless *assume* that

$$d^2\vec{\boldsymbol{r}} = 0, \tag{17.27}$$

but we will be careful not to assume that $d^2\vec{\boldsymbol{v}}$ is zero for other vectors—which, as we will see shortly, turns out to be false. Expanding (17.27), we have

$$0 = d^2\vec{\boldsymbol{r}} = d(d\vec{\boldsymbol{r}}) = d\left(\sigma^j\, \hat{\boldsymbol{e}}_j\right) = d\sigma^j\, \hat{\boldsymbol{e}}_j - \sigma^j \wedge d\hat{\boldsymbol{e}}_j \tag{17.28}$$

so that

$$0 = \hat{\boldsymbol{e}}_k \cdot d^2\vec{\boldsymbol{r}} = g_{kj}\, d\sigma^j - \sigma^j \wedge \omega_{kj}. \tag{17.29}$$

Rearranging terms and relabeling some indices, we have

$$0 = g_{ki}\, d\sigma^i + \omega_{kj} \wedge \sigma^j = g_{ki}\left(d\sigma^i + \omega^i{}_j \wedge \sigma^j\right). \tag{17.30}$$

Since the matrix (g_{ij}) is invertible, we can conclude that

$$0 = d\sigma^i + \omega^i{}_j \wedge \sigma^j. \tag{17.31}$$

A connection satisfying (17.31) is called *torsion free*, and a connection that is both torsion free and metric compatible is called a *Levi-Civita connection.*

We claim that, given $d\vec{\boldsymbol{r}}$, there is a unique Levi-Civita connection. For now, we give only a rough justification of this fact by counting the degrees of freedom. In n dimensions, a connection is determined by specifying the n^2 1-forms ω_{ij} in any basis. Metric compatibility (17.25) forces the n "diagonal" connection 1-forms ω_{ii} to be zero, and relates the remaining connection 1-forms in pairs, reducing the number of independent connection 1-forms to $n(n-1)/2$, each with n components. But the torsion-free condition consists of n 2-form equations, each with $n(n-1)/2$ components. Thus, the number of linear equations exactly matches the number of degrees of freedom, and we expect a unique solution.[1] A more complete derivation is given in Section 17.8.

17.7 Polar Coordinates III

Since the Levi-Civita connection is unique, the easiest way to determine it is often educated guesswork. We illustrate the technique using polar coordinates.

Our orthonormal basis of 1-forms is $\{dr, r\,d\phi\}$. Writing out the torsion-free condition, we have

$$d(dr) + \omega^r{}_r \wedge dr + \omega^r{}_\phi \wedge r\,d\phi = 0, \tag{17.32}$$

$$d(r\,d\phi) + \omega^\phi{}_r \wedge dr + \omega^\phi{}_\phi \wedge r\,d\phi = 0. \tag{17.33}$$

But metric compatibility tells us that

$$\omega^r{}_r = 0 = \omega^\phi{}_\phi, \tag{17.34}$$

$$\omega^\phi{}_r = -\omega^r{}_\phi \tag{17.35}$$

[1]This argument still holds when the left-hand sides of (17.25) and (17.31) are nonzero, but known. Thus, a connection is completely determined by specifying its "nonmetricity" and "torsion."

so that

$$\omega^r{}_\phi \wedge r\,d\phi = 0, \tag{17.36}$$

$$dr \wedge d\phi - \omega^r{}_\phi \wedge dr = 0. \tag{17.37}$$

In order to solve the second of these equations, $\omega^r{}_\phi$ must contain a term $-d\phi$. Guessing

$$\omega^r{}_\phi = -d\phi \tag{17.38}$$

solves not only the second equation but also the first, so we are done. A quick comparison with

$$\omega_{r\phi} = \hat{\boldsymbol{r}} \cdot d\hat{\boldsymbol{\phi}} = -d\phi \tag{17.39}$$

shows that our answer agrees with our computations in Section 17.1; there are no signs to worry about when "lowering an index" in this Euclidean example.

Yes, we could have solved the first equation explicitly, thus showing that there is no dr term in $\omega^r{}_\phi$, and in this particular example there's not much difference between these two solution strategies. But in more complicated examples, it is often most efficient to simply assume that certain components are zero; if it works, you're done.

17.8 UNIQUENESS OF THE LEVI-CIVITA CONNECTION

We now outline the derivation of an explicit formula for the connection 1-forms, thus also proving that, given $d\vec{\boldsymbol{r}}$, there is a unique Levi-Civita connection. We emphasize, however, that this formula is rarely the most efficient way to actually compute the connection 1-forms.

Introduce the components of the connection 1-forms by writing

$$\omega_{ij} = \Gamma_{ijk}\,\sigma^k. \tag{17.40}$$

The components Γ_{ijk} are known as *Christoffel symbols of the first kind.* Metric compatibility implies

$$0 = \omega_{ji} + \omega_{ij} = (\Gamma_{ijk} + \Gamma_{jik})\,\sigma^k \tag{17.41}$$

so that

$$\Gamma_{ijk} + \Gamma_{jik} = 0. \tag{17.42}$$

The torsion-free condition further implies that

$$0 = d\sigma_i + \omega_{ij} \wedge \sigma^j = d\sigma_i + \Gamma_{ijk}\,\sigma^k \wedge \sigma^j, \tag{17.43}$$

where we have introduced

$$\sigma_p = g_{pj}\sigma^j. \tag{17.44}$$

Note that

$$g(\sigma^i, \sigma_p) = \delta^i{}_p \tag{17.45}$$

since the sign in $g(\sigma^i, \sigma^j)$ cancels the one in the definition of σ_m. Thus, we also have

$$g(\sigma^i \wedge \sigma^j, \sigma_p \wedge \sigma_q) = \delta^i{}_p\delta^j{}_q - \delta^i{}_q\delta^j{}_p \tag{17.46}$$

using the definition of g on 2-forms, where we have used our new "up and down" index notation to keep track of the signs. Thus,

$$g(d\sigma_i, \sigma_p \wedge \sigma_q) = -\Gamma_{ijk}\, g(\sigma^k \wedge \sigma^j, \sigma_p \wedge \sigma_q) = \Gamma_{ipq} - \Gamma_{iqp}. \tag{17.47}$$

We can now find a formula for Γ_{ijk}, and hence for ω_{ij}, by writing

$$\begin{aligned}
2\Gamma_{ijk} &= \Gamma_{ijk} + \Gamma_{ijk} \\
&= \Gamma_{ijk} - \Gamma_{jik} \\
&= \Gamma_{ijk} + (-\Gamma_{ikj} + \Gamma_{ikj}) + (-\Gamma_{jki} + \Gamma_{jki}) - \Gamma_{jik} \\
&= (\Gamma_{ijk} - \Gamma_{ikj}) + (-\Gamma_{kij} + \Gamma_{kji}) + (\Gamma_{jki} - \Gamma_{jik}) \\
&= g(d\sigma_i, \sigma_j \wedge \sigma_k) - g(d\sigma_k, \sigma_i \wedge \sigma_j) + g(d\sigma_j, \sigma_k \wedge \sigma_i).
\end{aligned} \tag{17.48}$$

This construction will still work even if the left-hand sides of (17.41) and (17.43) are nonzero but known; the result is an expression for Γ_{ijk} that will contain those nonzero terms and is known as the *Koszul formula*. Thus, the "torsion" and "non-metricity" completely determine the connection, regardless of whether they are zero; the result for Levi-Civita connections is a special case.

17.9 TENSOR ALGEBRA

Tensors are multilinear maps on vectors. Differential forms are a type of tensor.

Consider the 1-form df for some function f. How does it act on a vector $\vec{\boldsymbol{v}}$? That's easy: by giving the directional derivative, namely

$$df(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{\nabla}} f \cdot \vec{\boldsymbol{v}}. \tag{17.49}$$

This is just the master formula in a new guise. Recall that the master formula tells us how to relate df and $\vec{\boldsymbol{\nabla}} f$, namely

$$df = \vec{\boldsymbol{\nabla}} f \cdot d\vec{\boldsymbol{r}} \tag{17.50}$$

and that a similar construction maps any vector field $\vec{\boldsymbol{F}}$ to a 1-form

$$F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}. \tag{17.51}$$

It is thus natural to define

$$(\vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}})(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{F}} \cdot \vec{\boldsymbol{v}}. \tag{17.52}$$

In an orthonormal basis we have

$$g_{ij}\, \sigma^j = \hat{\boldsymbol{e}}_i \cdot d\vec{\boldsymbol{r}} \tag{17.53}$$

so that

$$g_{ij}\, \sigma^j(\hat{\boldsymbol{e}}_k) = \hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_k = g_{ik}, \tag{17.54}$$

which forces

$$\sigma^j(\hat{\boldsymbol{e}}_k) = \delta^j{}_k. \tag{17.55}$$

Setting $\sigma_p = g_{pj}\sigma^j$ as before, we also have

$$g(\sigma^i, \sigma_p) = g_{pj}\, g(\sigma^i, \sigma^j) = g_{pj}\, g^{ij} = \delta^i{}_p = \sigma^i(\hat{\boldsymbol{e}}_p). \tag{17.56}$$

The action of higher-rank forms can be defined similarly. Recalling that

$$g(\alpha \wedge \beta, \sigma \wedge \tau) = g(\alpha, \sigma)\, g(\beta, \tau) - g(\alpha, \tau)\, g(\beta, \sigma), \tag{17.57}$$

we set

$$(\alpha \wedge \beta)(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}) = \alpha(\vec{\boldsymbol{v}})\, \beta(\vec{\boldsymbol{w}}) - \alpha(\vec{\boldsymbol{w}})\, \beta(\vec{\boldsymbol{v}}) \tag{17.58}$$

so that in particular

$$(\sigma^i \wedge \sigma^j)(\hat{\boldsymbol{e}}_p, \hat{\boldsymbol{e}}_q) = \sigma^i(\hat{\boldsymbol{e}}_p)\, \sigma^j(\hat{\boldsymbol{e}}_q) - \sigma^i(\hat{\boldsymbol{e}}_q)\, \sigma^j(\hat{\boldsymbol{e}}_p) = \delta^i{}_p\, \delta^j{}_q - \delta^i{}_q\, \delta^j{}_p. \tag{17.59}$$

17.10 COMMUTATORS

It is a remarkable fact that the expression

$$[\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}](f) = \vec{\boldsymbol{v}}\big(\vec{\boldsymbol{w}}(f)\big) - \vec{\boldsymbol{w}}\big(\vec{\boldsymbol{v}}(f)\big) \tag{17.60}$$

contains no second derivatives and therefore defines a new vector field $[\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}]$, which is called the *commutator* of the vector fields $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$.

Direct computation (which is easiest in a coordinate basis) now establishes the identity

$$d\alpha(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}) = \vec{\boldsymbol{v}}\big(\alpha(\vec{\boldsymbol{w}})\big) - \vec{\boldsymbol{w}}\big(\alpha(\vec{\boldsymbol{v}})\big) - \alpha([\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}]) \tag{17.61}$$

for any 1-form α, and applying this identity to an orthonormal basis results in

$$d\sigma^i(\hat{\boldsymbol{e}}_p, \hat{\boldsymbol{e}}_q) = 0 - 0 - \sigma^i([\hat{\boldsymbol{e}}_p, \hat{\boldsymbol{e}}_q]) \tag{17.62}$$

so that

$$g(d\sigma_i, \sigma_j \wedge \sigma_k) = d\sigma_i(\hat{\boldsymbol{e}}_j, \hat{\boldsymbol{e}}_k) = -\sigma_i([\hat{\boldsymbol{e}}_j, \hat{\boldsymbol{e}}_k]) = -\hat{\boldsymbol{e}}_i \cdot [\hat{\boldsymbol{e}}_j, \hat{\boldsymbol{e}}_k]. \tag{17.63}$$

We can use this expression to recover the standard formula for the (components of the) connection 1-forms in an orthonormal basis. We have

$$\begin{aligned} 2\Gamma_{ijk} &= g(d\sigma_i, \sigma_j \wedge \sigma_k) - g(d\sigma_k, \sigma_i \wedge \sigma_j) + g(d\sigma_j, \sigma_k \wedge \sigma_i) \\ &= -\hat{\boldsymbol{e}}_i \cdot [\hat{\boldsymbol{e}}_j, \hat{\boldsymbol{e}}_k] + \hat{\boldsymbol{e}}_k \cdot [\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_j] - \hat{\boldsymbol{e}}_j \cdot [\hat{\boldsymbol{e}}_k, \hat{\boldsymbol{e}}_i] \\ &= \hat{\boldsymbol{e}}_k \cdot [\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_j] + \hat{\boldsymbol{e}}_j \cdot [\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_k] - \hat{\boldsymbol{e}}_i \cdot [\hat{\boldsymbol{e}}_j, \hat{\boldsymbol{e}}_k], \end{aligned} \tag{17.64}$$

which is a special case of the *Koszul formula* (17.48) for the Levi-Civita connection. Thus, the connection 1-forms can be found by computing commutators of basis vectors.

This version of the Koszul formula simplifies even further in two dimensions, since i and j must be distinct and k must be either i or j. In either case, we obtain

$$2\Gamma_{ijk} = 2\hat{\boldsymbol{e}}_k \cdot [\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_j] \tag{17.65}$$

or equivalently

$$\omega_{ij} = \Gamma_{ijk}\, \sigma^k = \hat{\boldsymbol{e}}_k\, \sigma^k \cdot [\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_j] = [\hat{\boldsymbol{e}}_i, \hat{\boldsymbol{e}}_j] \cdot d\vec{\boldsymbol{r}}. \tag{17.66}$$

17.11 PROBLEMS

1. **Spherical Coordinates II**

   Consider spherical coordinates (r,θ,ϕ) and the adapted orthonormal basis

   $$\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \hat{\boldsymbol{e}}_3\} = \{\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}\}.$$

   The "infinitesimal displacement vector" $d\vec{\boldsymbol{r}}$ relates this basis to an orthonormal basis of 1-forms via

   $$d\vec{\boldsymbol{r}} = dr\, \hat{\boldsymbol{r}} + r\, d\theta\, \hat{\boldsymbol{\theta}} + r \sin\theta\, d\phi\, \hat{\boldsymbol{\phi}}.$$

   Both sides of this equation are really *vector valued 1-forms.*

(a) Determine the exterior derivative of each basis vector (not 1-form) above; that is, compute $d\hat{\boldsymbol{r}}$, $d\hat{\boldsymbol{\theta}}$, and $d\hat{\boldsymbol{\phi}}$.
What sort of a beast should you get?
What is the position vector $\vec{\boldsymbol{r}}$ in this basis?

(b) Compute $\omega_{ij} = \hat{\boldsymbol{e}}_i \cdot d\hat{\boldsymbol{e}}_j$ for $i, j = 1, 2, 3$.
What sort of a beast should you get?

(c) Compute $\Omega_{ij} = d\omega_{ij} + \omega_{ik} \wedge \omega_{kj}$ for $i, j = 1, 2, 3$ (and where there is an implicit sum over k).
What sort of a beast should you get?

2. **Spherical Coordinates III**

Consider the sphere of radius r in spherical coordinates (θ,ϕ), with line element
$$ds^2 = r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2\right) .$$

(a) Find the connection 1-forms ω_{ij} in this basis.

(b) Compute $\Omega_{ij} = d\omega_{ij} + \omega_{ik} \wedge \omega_{kj}$ for $i, j = 1, 2$ (and where there is an implicit sum over k).

(c) Compare your answers (and your computations) with those from the previous problem.

≻ Chapter 18 ≺

Curvature

18.1 Curves

How much does a curve bend? Take the circle that best fits the curve at a given point, and use the radius of this circle to specify the curvature of the curve at that point. Since large circles have *less* curvature than small circles, define the curvature to be

$$\kappa = \frac{1}{r}, \tag{18.1}$$

where r is the radius of the best-fit circle. The *unit tangent vector* to a curve is given by

$$\hat{\boldsymbol{T}} = \frac{d\vec{\boldsymbol{r}}}{|d\vec{\boldsymbol{r}}|} = \frac{d\vec{\boldsymbol{r}}}{ds}, \tag{18.2}$$

where, as usual, s denotes arclength along the curve. For a circle, it is easy to see that

$$\frac{d\hat{\boldsymbol{T}}}{ds} = \frac{1}{r}\,\hat{\boldsymbol{N}}, \tag{18.3}$$

where $\hat{\boldsymbol{N}}$ is the *inward*-pointing normal vector. Since this computation depends only on the first derivative of the tangent vector, the result must be the same for the best-fit circle and for the curve itself. We have therefore shown that

$$\frac{d\hat{\boldsymbol{T}}}{ds} = \kappa\hat{\boldsymbol{N}}, \tag{18.4}$$

where $\hat{\boldsymbol{N}}$ is the *principal unit normal vector*, which points in the direction of bending.

This construction is depicted in Figure 18.1, which shows the best-fit circle at two points on a parabola, together with vectors pointing to the center of each circle. The magnitude of the arrows gives the *radius of curvature* at each point, which is the reciprocal of the curvature, and the direction of the arrows is the principal unit normal vector at that point.

Rewriting (18.4) somewhat, we obtain

$$\kappa\, ds = d\hat{\boldsymbol{T}} \cdot \hat{\boldsymbol{N}} = -d\hat{\boldsymbol{N}} \cdot \hat{\boldsymbol{T}}, \tag{18.5}$$

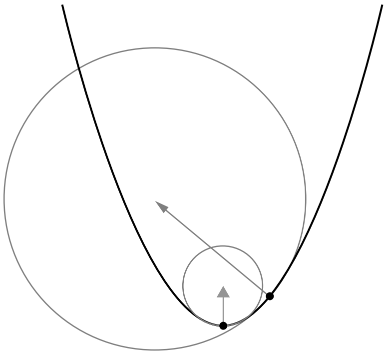


FIGURE 18.1. The best-fit circle at two points along a parabola.

where the last equality follows since $\hat{\boldsymbol{N}} \perp \hat{\boldsymbol{T}}$.[1] Thus, the curvature of a curve is a measure of the change in the normal vector along the curve.

18.2 SURFACES

One way to describe a surface is as the level surface of some function, which can be taken to be a coordinate. In this section, we will work in three Euclidean dimensions with coordinates (x^1,x^2,x^3) and assume that our surface is given as $x^3 =$ constant. Thus, on our surface we have $dx^3 = 0$. More generally, in orthogonal coordinates, we can assume that $\sigma^3 = 0$ on our surface.

What is the curvature? Reasoning by analogy, we consider the curvature of coordinate curves in the surface. Introduce an orthonormal basis $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \hat{\boldsymbol{n}} = \hat{\boldsymbol{e}}_3\}$ and look at how the normal vector $\hat{\boldsymbol{n}}$ changes along such curves. We have

$$-d\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{e}}_i = \hat{\boldsymbol{n}} \cdot d\hat{\boldsymbol{e}}_i = \omega_{3i}, \tag{18.6}$$

so (some of the) connection 1-forms appear to be telling us something about curvature!

Recall now the structure equations, one of which says that

$$d\sigma^3 + \omega^3{}_i \wedge \sigma^i = 0. \tag{18.7}$$

[1]The derivative of a unit vector must be perpendicular to the original vector.

But *on the surface* we can assume that $\sigma^3 = 0$ everywhere, so that $d\sigma^3 = 0$, and we have

$$\omega^3{}_1 \wedge \sigma^1 + \omega^3{}_2 \wedge \sigma^2 = 0. \tag{18.8}$$

If we write

$$\omega^3{}_i = \Gamma^3{}_{ij}\,\sigma^j, \tag{18.9}$$

then the structure equation forces $\Gamma^3{}_{12} = \Gamma^3{}_{21}$ on the surface. The 2×2 matrix

$$S = -\left(\Gamma^3{}_{ij}\right) \tag{18.10}$$

(evaluated on the surface, and with $i, j = 1, 2$) is therefore symmetric and is called the *shape operator*.

The shape operator S is, of course, dependent on the basis used to compute it. However, its trace and determinant are not. The eigenvalues of S, usually denoted κ_1 and κ_2, are called the *principal curvatures* of the surface and give the maximum and minimum curvatures for any curve through the given point. Since S is symmetric, these eigenvalues correspond to orthogonal eigenvectors, and the directions of maximum and minimum curvature are perpendicular. The trace $\kappa_1 + \kappa_2$ of S is called the *mean curvature*, and the determinant $K = \kappa_1\kappa_2$ of S is called the *Gaussian curvature*. As we will see in Sections 18.3 and 18.5, only the Gaussian curvature is an intrinsic property of the surface.

18.3 EXAMPLES IN THREE DIMENSIONS

Consider first the xy-plane. As shown in Figure 18.2, we have

$$\hat{\boldsymbol{e}}_1 = \hat{\boldsymbol{x}}, \tag{18.11}$$
$$\hat{\boldsymbol{e}}_2 = \hat{\boldsymbol{y}}, \tag{18.12}$$
$$\hat{\boldsymbol{n}} = \hat{\boldsymbol{e}}_3 = \hat{\boldsymbol{z}}, \tag{18.13}$$

and of course

$$d\hat{\boldsymbol{n}} = 0 \tag{18.14}$$

so that S is the zero matrix and all of the curvatures are zero.

What about a cylinder? As shown in Figure 18.3, now we have

$$\hat{\boldsymbol{e}}_1 = \hat{\boldsymbol{\phi}}, \tag{18.15}$$
$$\hat{\boldsymbol{e}}_2 = \hat{\boldsymbol{z}}, \tag{18.16}$$
$$\hat{\boldsymbol{n}} = \hat{\boldsymbol{e}}_3 = \hat{\boldsymbol{r}}, \tag{18.17}$$

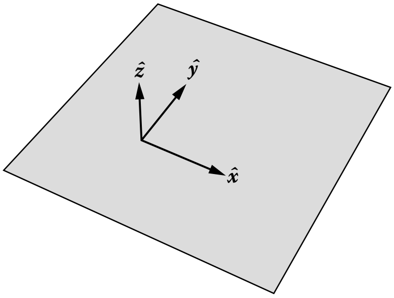


FIGURE 18.2. The xy-plane.

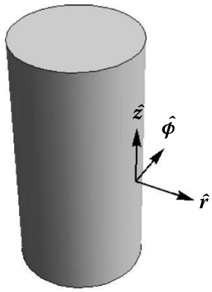


FIGURE 18.3. A cylinder.

and now

$$d\hat{\boldsymbol{n}} = d\phi\,\hat{\boldsymbol{\phi}} \tag{18.18}$$

so that

$$S = \begin{pmatrix} 1/r & 0 \\ 0 & 0 \end{pmatrix}. \tag{18.19}$$

Thus, the principal curvatures are $\kappa_1 = 0$ and $\kappa_2 = 1/r$. This result should not be a surprise—the cylinder is curved only in one direction. Furthermore, the Gaussian curvature vanishes ($\kappa_1\kappa_2 = 0$); the cylinder is indeed flat and can be made by rolling up a piece of paper, without stretching or tearing.

What about a sphere? As shown in Figure 18.4, in this case we have

$$\hat{\boldsymbol{e}}_1 = \hat{\boldsymbol{\theta}}, \tag{18.20}$$

$$\hat{\boldsymbol{e}}_2 = \hat{\boldsymbol{\phi}}, \tag{18.21}$$

$$\hat{\boldsymbol{n}} = \hat{\boldsymbol{e}}_3 = \hat{\boldsymbol{r}}, \tag{18.22}$$

and now

$$d\hat{\boldsymbol{n}} = d\theta\,\hat{\boldsymbol{\theta}} + \sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}, \tag{18.23}$$

which leads to

$$S = \begin{pmatrix} 1/r & 0 \\ 0 & 1/r \end{pmatrix} \tag{18.24}$$

so that the principal curvatures κ_1 and κ_2 are both $1/r$. Again, this result should not be a surprise—the sphere has the same curvature in all directions. Furthermore, in this case the Gaussian curvature does not vanish ($\kappa_1\kappa_2 = 1/r^2$); the sphere cannot be made from a piece of paper without stretching or tearing.

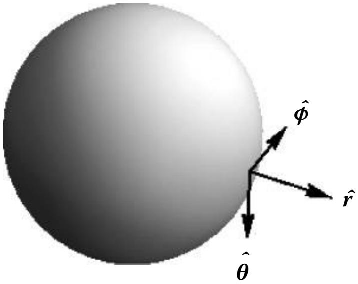


FIGURE 18.4. A sphere.

These examples lead us to suspect that only the Gaussian curvature is an *intrinsic* property of a surface. This conjecture turns out to be correct, as we will see shortly. All of the other curvatures—including the curvature of a curve—are *extrinsic*; they are a property of the way the surface (or curve) sits inside $\mathbb{R}^3$, but cannot be detected by an ant walking along the surface (or curve).

18.4 CURVATURE

In Section 17.6, we computed

$$\begin{aligned} d^2\vec{\boldsymbol{r}} &= d\left(\hat{\boldsymbol{e}}_j\, \sigma^j\right) \\ &= d\hat{\boldsymbol{e}}_j \wedge \sigma^j + \hat{\boldsymbol{e}}_j\, d\sigma^j \\ &= \hat{\boldsymbol{e}}_i \left(\omega^i{}_j \wedge \sigma^j + d\sigma^i\right). \end{aligned} \tag{18.25}$$

The vector components of this equation are called the *first structure equation*, which takes the form

$$\Theta^i = \omega^i{}_j \wedge \sigma^j + d\sigma^i, \tag{18.26}$$

where the Θ^i are the *torsion 2-forms*. Since we are assuming that

$$d^2\vec{\boldsymbol{r}} = 0, \tag{18.27}$$

the torsion vanishes and the structure equation reduces to

$$0 = d\sigma^i + \omega^i{}_j \wedge \sigma^j. \tag{18.28}$$

We can do a similar computation starting with $d\hat{\boldsymbol{e}}_j$ instead of $d\vec{\boldsymbol{r}}$. We have

$$\begin{aligned} d^2\hat{\boldsymbol{e}}_j &= d\left(\hat{\boldsymbol{e}}_k\,\omega^k{}_j\right) \\ &= d\hat{\boldsymbol{e}}_k \wedge \omega^k{}_j + \hat{\boldsymbol{e}}_k\,d\omega^k{}_j \\ &= \hat{\boldsymbol{e}}_i\left(\omega^i{}_k \wedge \omega^k{}_j + d\omega^i{}_j\right). \end{aligned} \tag{18.29}$$

The components of this equation are the *second structure equation*, which takes the form

$$\Omega^i{}_j = d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j, \tag{18.30}$$

where the $\Omega^i{}_j$ are the *curvature 2-forms*. We *cannot*, in general, assume that the curvature 2-forms vanish. However, the curvature 2-forms $\Omega^i{}_j$ are not all independent, as we now show.

Consider first an orthonormal basis, in which metric compatibility takes the form

$$\omega_{ij} + \omega_{ji} = 0, \tag{18.31}$$

which clearly implies that

$$d\omega_{ij} + d\omega_{ji} = 0. \tag{18.32}$$

What about the second term in the curvature 2-forms? We have

$$\omega_{ik} \wedge \omega^k{}_j = -\omega^k{}_j \wedge \omega_{ik} = \omega^k{}_j \wedge \omega_{ki} = \omega_{kj} \wedge \omega^k{}_i = -\omega_{jk} \wedge \omega^k{}_i. \tag{18.33}$$

Putting these computations together, we have shown that, in an orthonormal basis, we have

$$\Omega_{ij} = -\Omega_{ji} \tag{18.34}$$

since

$$\Omega_{ij} = d\omega_{ij} + \omega_{ik} \wedge \omega^k{}_j, \tag{18.35}$$

which is antisymmetric in i and j.

The argument leading to (18.34) used properties of an orthonormal basis, but the result is true in any basis. To see this, we repeat the above computation in a general basis. Metric compatibility now takes the form

$$\omega_{ij} + \omega_{ji} = dg_{ij}, \tag{18.36}$$

which still leads to (18.32), since $d^2 g_{ij} = 0$. What about the second term in (18.35)? We must carefully lower the index, yielding[2]

$$\Omega_{ij} = g_{ik}\Omega^k{}_j = g_{ik}d\omega^k{}_j + g_{ik}\omega^k{}_l \wedge \omega^l{}_j = g_{ik}d\omega^k{}_j + \omega_{ik} \wedge \omega^k{}_j. \tag{18.37}$$

[2]We can no longer move the factor of g_{ik} inside the exterior derivative, since it is not necessarily constant.

Using the product rule on the first term leads to

$$\begin{aligned}\Omega_{ij} &= d(g_{ik}\omega^k{}_j) - dg_{ik} \wedge \omega^k{}_j + \omega_{ik} \wedge \omega^k{}_j \\ &= d\omega_{ij} + (\omega_{ik} - dg_{ik}) \wedge \omega^k{}_j = d\omega_{ij} - \omega_{ki} \wedge \omega^k{}_j,\end{aligned} \tag{18.38}$$

which agrees with (18.35) in the orthonormal case. Each of these terms is clearly antisymmetric in i and j, thus establishing (18.34) in the general case.

Thus, of the n^2 possible curvature 2-forms (with lowered indices, that is, Ω_{ij}) in n-dimensions, n are identically zero, and the remaining $n^2 - n$ curvature 2-forms are equal and opposite in pairs. Thus, there are only $\binom{n}{2} = \frac{1}{2}n(n-1)$ independent curvature 2-forms.

18.5 CURVATURE IN THREE DIMENSIONS

In three Euclidean dimensions, the curvature 2-forms vanish. Thus,

$$0 = \Omega^i{}_j = d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j \tag{18.39}$$

and in particular

$$0 = \Omega^1{}_2 = d\omega^1{}_2 + \omega^1{}_k \wedge \omega^k{}_2 = d\omega^1{}_2 + \omega^1{}_3 \wedge \omega^3{}_2. \tag{18.40}$$

If we were instead to look only at the two-dimensional surface where x^3 is constant, we would compute

$$\tilde{\Omega}^1{}_2 = d\omega^1{}_2 + \omega^1{}_k \wedge \omega^k{}_2 = d\omega^1{}_2 \neq 0 \tag{18.41}$$

since now $k = 1, 2$; we have written $\tilde{\Omega}^i{}_j$ for the two-dimensional curvature 2-forms to distinguish them from the three-dimensional curvature 2-forms $\Omega^i{}_j$.

However, it is important to realize that no tilde is needed on $\omega^1{}_2$. To see this, recall the structure equation

$$d\sigma^i + \omega^i{}_j \wedge \sigma^j = 0. \tag{18.42}$$

The two-dimensional version of this equation is obtained from the three-dimensional version simply by setting x^3 = constant and $\sigma^3 = 0$. Thus, the same is true for the solutions, and the two-dimensional $\omega^1{}_2$ can be similarly obtained from the three-dimensional $\omega^1{}_2$.[3]

[3] Why does this argument fail for $\Omega^1{}_2$? Because the three-dimensional version contains an extra term, $\omega^1{}_3 \wedge \omega^3{}_2$, which simply isn't present in the two-dimensional version. This situation does not occur for $\omega^1{}_2$, where setting $\sigma^3 = 0$ eliminates all terms containing the index "3."

But we already know $\omega^1{}_2$, since on the surface we must have

$$\begin{aligned} d\omega^1{}_2 &= -\omega^1{}_3 \wedge \omega^3{}_2 \\ &= \omega^3{}_1 \wedge \omega^3{}_2 \\ &= \Gamma^3{}_{1k}\,\sigma^k \wedge \Gamma^3{}_{2l}\,\sigma^l \\ &= K\,\sigma^1 \wedge \sigma^2, \end{aligned} \tag{18.43}$$

where

$$K = \det(-S) = \det(S) \tag{18.44}$$

is the Gaussian curvature. Thus, for a two-dimensional surface, we have

$$\Omega^1{}_2 = d\omega^1{}_2 = K\,\omega, \tag{18.45}$$

where we have dropped the tilde.

This remarkable result is due to Gauss and is known as the *Theorema Egregium* ("Outrageous Theorem"): The Gaussian curvature, originally defined extrinsically, is in fact an *intrinsic* property of the surface.

18.6 COMPONENTS

Just as the components of the connection 1-forms are the Christoffel symbols, related by

$$\omega^i{}_j = \Gamma^i{}_{jk}\,\sigma^k, \tag{18.46}$$

the components of curvature also have names.[4] We write

$$\Omega^i j = \frac{1}{2} R^i{}_{jkl}\,\sigma^k \wedge \sigma^l, \tag{18.47}$$

where the $R^i{}_{jkl}$ are the components of the *Riemann curvature tensor*. However, (18.47) alone determines only the combinations $R^i{}_{jkl} - R^i{}_{jlk}$, so we assume by convention that

$$R^i{}_{jlk} = -R^i{}_{jkl}. \tag{18.48}$$

The *Ricci curvature tensor* is defined as the *trace* of the Riemann tensor, defined by

$$\Omega^k{}_i(\hat{\boldsymbol{e}}_k) = R^m{}_{imj}\,\sigma^j = R_{ij}\,\sigma^j, \tag{18.49}$$

[4]Similarly, the components of the torsion 2-forms Θ^i are the components $T^i{}_{jk}$ of the *torsion tensor*, where $\Theta^i = \frac{1}{2} T^i{}_{jk}\,\sigma^j \wedge \sigma^k$.

where the last expression defines the Ricci components R_{ij}. Equivalently, we have

$$g(\Omega^k{}_i, \sigma_k \wedge \sigma_j) = R_{ij}. \tag{18.50}$$

Finally, the *Ricci curvature scalar* is the trace of the Ricci tensor, defined by

$$R = R_{ij}\, g^{ij} = g(\sigma^i, R_{ij}\, \sigma^j). \tag{18.51}$$

One way to visualize a 4-index object such as $R^m{}_{inj}$ is to note that if i and j are fixed, then this expression corresponds to a 4×4 matrix of coefficients. Thus, for each pair (i, j), we have such a 4×4 matrix. But i, j themselves denote the components of a 4×4 matrix, so $R^m{}_{inj}$ can be thought of as a 4×4 matrix (with components labeled by (i, j)) of 4×4 matrices (with components labeled by (m, n)).

If we take the trace of each of the "inner" 4×4 matrices, we replace each one with a number, namely its trace. The result? An ordinary 4×4 matrix, with components R_{ij}—and trace R.

18.7 BIANCHI IDENTITIES

Since σ^i is an (ordinary) differential form, we must have

$$\begin{aligned} 0 = -d^2\sigma^i &= d\left(\omega^i{}_k \wedge \sigma^k\right) \\ &= d\omega^i{}_k \wedge \sigma^k - \omega^i{}_k \wedge d\sigma^k \\ &= d\omega^i{}_j \wedge \sigma^j + \omega^i{}_k \wedge \omega^k{}_j \wedge \sigma^j \\ &= \left(d\omega^i{}_j + \omega^i{}_k \wedge \omega^k{}_j\right) \wedge \sigma^k \\ &= \Omega^i{}_j \wedge \sigma^j, \end{aligned} \tag{18.52}$$

which is the *first Bianchi identity.* A similar argument starting with the connection 1-forms yields

$$\begin{aligned} 0 = d^2\omega^i{}_j &= d\left(\Omega^i{}_j - \omega^i{}_k \wedge \omega^k{}_j\right) \\ &= d\Omega^i{}_j - d\omega^i{}_k \wedge \omega^k{}_j + \omega^i{}_k \wedge d\omega^k{}_j \\ &= d\Omega^i{}_j - \left(\Omega^i{}_k - \omega^i{}_l \wedge \omega^l{}_k\right) \wedge \omega^k{}_j + \omega^i{}_k \wedge \left(\Omega^k{}_j - \omega^k{}_l \wedge \omega^l{}_j\right) \\ &= d\Omega^i{}_j + \omega^i{}_k \wedge \Omega^k{}_j - \Omega^i{}_k \wedge \omega^k{}_j \end{aligned} \tag{18.53}$$

since the remaining terms cancel by suitable relabeling of indices; this is the *second Bianchi identity*, which is often simply called "the" Bianchi identity.

In terms of components, the first Bianchi identity can be written in the form

$$R^i{}_{jkl}\,\sigma^k \wedge \sigma^l \wedge \sigma^j = 0, \tag{18.54}$$

which implies

$$R^i{}_{jkl} + R^i{}_{klj} + R^i{}_{ljk} = 0 \tag{18.55}$$

(since the components $R^i{}_{jkl}$ are antisymmetric in their last two indices). The second Bianchi identity involves derivatives of these components and can be interpreted as the vanishing of the "covariant curl" of the Riemann tensor.

The Bianchi identities imply some further symmetries on the components of the Riemann and Ricci tensors. Lowering an index for convenience, the antisymmetry of 2-forms leads to

$$R_{ijlk} = -R_{ijkl} \tag{18.56}$$

by convention, and metric compatibility leads to

$$R_{jikl} = -R_{ijkl} \tag{18.57}$$

as shown in Section 18.4. We can now show that the components of the Ricci tensor are symmetric, since

$$R_{ij} = R^m{}_{imj} = -R^m{}_{mji} - R^m{}_{jim} = R^m{}_{jmi} = R_{ji}. \tag{18.58}$$

Finally, we compute

$$\begin{aligned} R_{ijkl} &= -R_{ijlk} = R_{jilk} \\ &= -R_{jlki} - R_{jkil} \\ &= R_{ljki} + R_{kjil} \\ &= -R_{lkij} - R_{lijk} - R_{kilj} - R_{klji} \\ &= 2R_{klij} + R_{iljk} + R_{iklj} \\ &= 2R_{klij} - R_{ijkl} \end{aligned} \tag{18.59}$$

so that

$$R_{ijkl} = R_{klij}. \tag{18.60}$$

18.8 GEODESIC CURVATURE

Consider a curve C in a two-dimensional surface Σ. We will assume Euclidean signature. You can imagine that Σ sits inside Euclidean $\mathbb{R}^3$ if

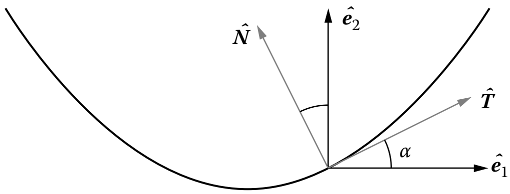


FIGURE 18.5. The tangent and normal vectors to a curve.

desired, but this is not necessary. Choose an orthonormal basis of vectors $\{\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2\}$ as usual. Then, as shown in Figure 18.5, the unit tangent vector $\hat{\boldsymbol{T}}$ to the curve must satisfy

$$\hat{\boldsymbol{T}} = \cos\alpha\,\hat{\boldsymbol{e}}_1 + \sin\alpha\,\hat{\boldsymbol{e}}_2 \tag{18.61}$$

for some angle α. Differentiating, we obtain

$$\begin{aligned} d\hat{\boldsymbol{T}} &= (-\sin\alpha\,\hat{\boldsymbol{e}}_1 + \cos\alpha\,\hat{\boldsymbol{e}}_2)\;d\alpha + \cos\alpha\,\omega^2{}_1\,\hat{\boldsymbol{e}}_2 + \sin\alpha\,\omega^1{}_2\,\hat{\boldsymbol{e}}_1 \\ &= (-\sin\alpha\,\hat{\boldsymbol{e}}_1 + \cos\alpha\,\hat{\boldsymbol{e}}_2)\left(d\alpha - \omega^1{}_2\right) \end{aligned} \tag{18.62}$$

using the symmetry of ω_{ij} and the fact that in Euclidean signature we can raise and lower indices freely. The vector in parentheses is a unit vector $\hat{\boldsymbol{N}}$ in Σ that is orthogonal to the curve, as shown in Figure 18.5. That is,

$$\hat{\boldsymbol{N}} = -\sin\alpha\,\hat{\boldsymbol{e}}_1 + \cos\alpha\,\hat{\boldsymbol{e}}_2, \tag{18.63}$$

which can be thought of as $\hat{\boldsymbol{e}}_3 \times \hat{\boldsymbol{T}}$, where $\hat{\boldsymbol{e}}_3$ is the normal vector to Σ in $\mathbb{R}^3$.[5] The curvature associated with $\hat{\boldsymbol{T}}$ and $\hat{\boldsymbol{N}}$ is called the *geodesic curvature* κ_g of C and is given by

$$\kappa_g\,ds = d\hat{\boldsymbol{T}}\cdot\hat{\boldsymbol{N}} = d\alpha - \omega^1{}_2. \tag{18.64}$$

A curve with $\kappa_g = 0$ is called a *geodesic*. It is clear from the definition that the geodesics in the plane are straight lines. What are the geodesics on a sphere? A curve on a sphere that does not bend to the left or right is surely a circle, and we can assume without loss of generality that it is a circle of constant latitude. (Why?) The geodesic curvature therefore vanishes if

$$0 = d\hat{\boldsymbol{\phi}}\cdot(-\hat{\boldsymbol{\theta}}) = \cos\theta\,d\phi, \tag{18.65}$$

[5] Do not confuse $\hat{\boldsymbol{N}}$ as used here with the principal unit normal vector $\hat{\boldsymbol{N}}$ in Section 18.1, which depends only on the curve. The normal vector $\hat{\boldsymbol{N}}$ defined by (18.63) depends on the orientation of the *surface*, not the direction of bending of the curve.

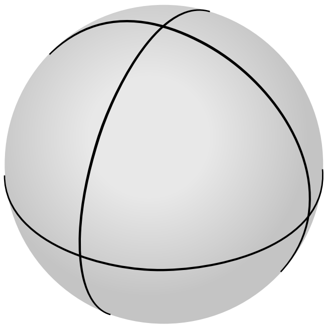


FIGURE 18.6. Some geodesics on a sphere.

which is true only at the equator, where $\theta = \pi/2$. Thus, the geodesics on the sphere are *great circles*, as shown in Figure 18.6.

18.9 GEODESIC TRIANGLES

What is the total curvature around a closed curve? Using the definition in Section 18.8, if we integrate the geodesic curvature around any closed curve we have

$$\oint \kappa_g \, ds = \oint d\alpha - \oint \omega^1{}_2 = 2\pi - \oint \omega^1{}_2 \tag{18.66}$$

since $\hat{\boldsymbol{T}}$ must return to its original orientation. We can now use Stokes' Theorem on the last term, which tells us that

$$\oint \omega^1{}_2 = \int d\omega^1{}_2, \tag{18.67}$$

where the integral on the right is over the *interior* of the curve. Furthermore, recall that

$$\Omega^1{}_2 = d\omega^1{}_2 = K\,\omega, \tag{18.68}$$

where K is the Gaussian curvature and ω the orientation of Σ. Putting this all together, we obtain

$$\int K\,\omega + \oint \kappa_g \, ds = 2\pi. \tag{18.69}$$

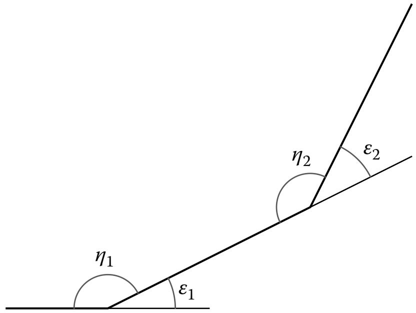


FIGURE 18.7. Part of a piecewise smooth closed curve, showing the exterior (ϵ_i) and interior (η_i) angles. (The sides do not need to be straight.)

What if the curve is only piecewise smooth? Then the integral around the curve will not pick up the discrete jumps in α at the corners. In this case (18.69) becomes

$$\int K\,\omega + \oint \kappa_g\,ds = 2\pi - \sum \epsilon_i = 2\pi - \sum(\pi - \eta_i), \tag{18.70}$$

where ϵ_i is the *exterior* angle at the ith corner, namely the angle through which the tangent vector rotates at that corner, and η_i is the corresponding *interior* angle, as shown in Figure 18.7.

An immediate application of (18.70) is a formula for geodesic triangles on a sphere, for which $\kappa_g = 0$ and $\int K\,\omega$ gives the area of the triangle divided by the square of the radius a. Thus, the area A of a geodesic triangle on a sphere is given by

$$\frac{A}{a^2} = 2\pi - \sum(\pi - \eta_i) = \sum \eta_i - \pi, \tag{18.71}$$

where the sum runs from $i = 1$ to $i = 3$. The area of such a triangle is therefore proportional to the *angle excess* of the triangle, that is, the amount by which the sum of the interior angles exceeds π. This argument also shows that the sum of the angles in a geodesic triangle on a sphere *always* exceeds π. An example is shown in Figure 18.8.

In the plane, $K = 0$, and the same argument reduces to the fact that the angle sum of a triangle in the plane is always π.

Spacelike hyperboloids $x^2 - t^2 = -a^2$ in (two-dimensional) Minkowski space provide a model for *hyperbolic geometry*, in which the Gaussian curvature is constant but negative. (See Figure 9.4, and see Chapter 14 of [2]

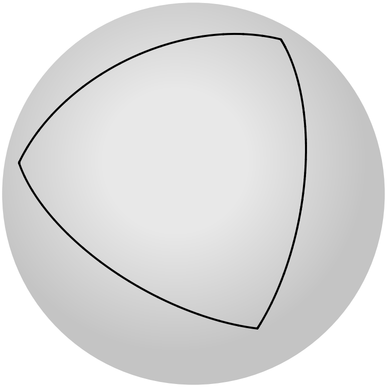


FIGURE 18.8. A geodesic triangle on a sphere, with three right angles.

for further details.) A similar argument to that given above for spheres leads again to (18.71), but with a minus sign on the left-hand side; the area of hyperbolic triangles is proportional to the *angle deficit*.

18.10 THE GAUSS–BONNET THEOREM

Consider now an (oriented) compact surface $\Sigma \in \mathbb{R}^3$. Any such surface has a *rectangular decomposition*; that is, it can be covered by finitely many quadrilaterals, the sides of which are smooth curves, with the further assumption that any two such quadrilaterals overlap, if at all, either in a single common vertex or in a single common edge.

An example is given in Figure 18.9, which shows a single quadrilateral on the sphere, whose sides are *not* all geodesics. Arranging four such quadrilaterals next to each other around the sphere results in a quadrilateral cover of the sphere, although in this particular example the extra quadrilateral "caps" at the top and bottom would be circles.

Just as with the triangles in Section 18.9, we can consider the total curvature around one of these quadrilaterals. Since we are not assuming that the curves are geodesics, we have

$$\int K\,\omega + \oint \kappa_g\,ds = 2\pi - \sum(\pi - \eta_i) = \sum \eta_i - 2\pi \tag{18.72}$$

since the sum now runs from $i = 1$ to $i = 4$. If we add up these expressions over the entire surface, that is, over all of our quadrilaterals, several things

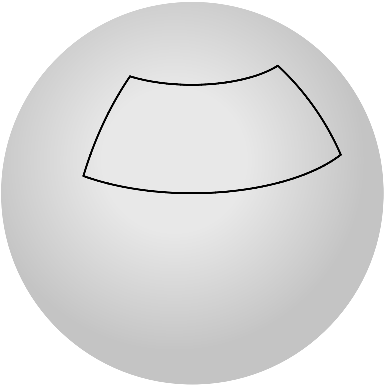


FIGURE 18.9. A (non-geodesic) quadrilateral on the sphere.

happen. First of all, $\kappa_g\,ds$ is being integrated twice over each edge, with opposite orientations. These integrals cancel, and the κ_g term does not contribute overall. Furthermore, the sum over the interior angles η_i is now a sum over *all* interior angles, which must add up to 2π at each vertex. This term thus becomes $2\pi v$, where v is the number of vertices. Finally, there is a factor of 2π for each quadrilateral, that is, for each "face," which contributes $2\pi f$, where f is the number of faces. We have therefore shown that

$$\int_\Sigma K\,\omega = 2\pi v - 2\pi f. \tag{18.73}$$

There is one further simplification. Since we assumed that each face is a quadrilateral, each face has four edges. However, each edge belongs to two faces, so that

$$e = \frac{4\,f}{2}, \tag{18.74}$$

where e is the number of edges, which in turn implies that

$$e - f = f. \tag{18.75}$$

Putting this all together, we obtain the *Gauss–Bonnet Theorem*, which says that

$$\int_\Sigma K\,\omega = 2\pi(v - e + f) = 2\pi\chi(\Sigma), \tag{18.76}$$

where the last equality defines the *Euler characteristic* $\chi(\Sigma)$ of Σ.

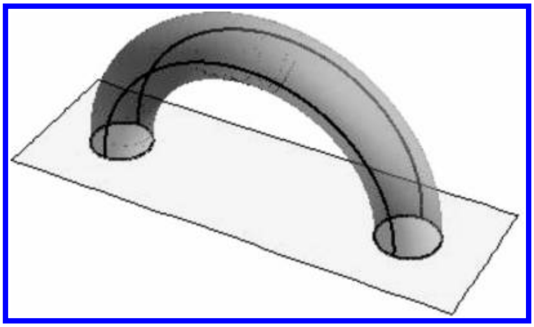


FIGURE 18.10. Adding a handle to a surface.

Although we have only established the Gauss–Bonnet Theorem for decompositions in terms of quadrilaterals, it is a remarkable fact that the Euler characteristic does not depend on whether quadrilaterals or arbitrary polygons are used. In other words, the Euler characteristic is a topological invariant of the surface, and we can compute it with any polygons we choose, including quadrilaterals.

The Gauss–Bonnet Theorem is even more remarkable than the *Theorema Egregium*! The Euler characteristic is a purely topological property, whereas the Gaussian curvature is purely geometric. The latter requires both a notion of distance and differentiability; the former does not. Nonetheless, they are related to each other.

In the example shown in Figure 18.9, extended as described above to cover the sphere with 4 quadrilaterals and 2 caps, there are 8 vertices, 12 edges, and 6 faces, so

$$\chi(\mathbb{S}^2) = 8 - 12 + 6 = 2 \tag{18.77}$$

on the sphere.

Another example demonstrates the power of the Gauss–Bonnet Theorem. As just computed, the Euler characteristic of a sphere is 2, and it is easily seen that (18.76) is satisfied. Adding a handle decreases the Euler characteristic by 2, as shown in Figure 18.10: Two faces (the circles) have been removed; two faces (the top and bottom of the tube) and two edges (where the faces meet) have been added. Adding a handle to a sphere yields a torus, whose Euler characteristic turns out to be 0. Thus, the total Gaussian curvature of a torus must be 0. This is plausible, since the inner half of a donut curves the opposite way from the outer half. One conse-

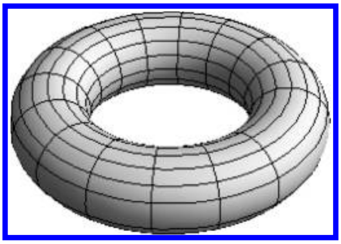


FIGURE 18.11. A torus in $\mathbb{R}^3$.

quence of the Gauss–Bonnet Theorem is that it is not possible to construct a torus so that the Gaussian curvature is everywhere positive!

18.11 THE TORUS

The torus $\mathbb{T}$, shown in Figure 18.11, can be parametrized as a surface of revolution about the z-axis, as shown in Figure 18.12, resulting in

$$x = (R + r\,\cos\theta)\,\cos\phi, \tag{18.78}$$

$$y = (R + r\,\cos\theta)\,\sin\phi, \tag{18.79}$$

$$z = r\,\sin\theta, \tag{18.80}$$

from which it follows by direct computation that

$$ds^2 = r^2\,d\theta^2 + (R + r\,\cos\theta)^2\,d\phi^2. \tag{18.81}$$

Alternatively, (18.81) follows directly from the Pythagorean theorem, using arclength to compute small distances in the θ and ϕ directions.

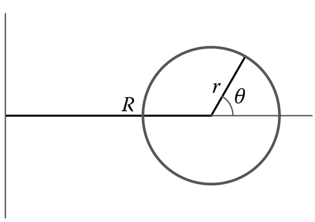


FIGURE 18.12. Parametrizing the torus as a surface of revolution about the z-axis.

Our orthonormal basis of 1-forms is therefore

$$\sigma^\theta = r\,d\theta, \tag{18.82}$$
$$\sigma^\phi = (R + r\,\cos\theta)\,d\phi. \tag{18.83}$$

The structure equations are now

$$-\omega^\theta{}_\phi \wedge \sigma^\phi = d\sigma^\theta = 0, \tag{18.84}$$
$$-\omega^\phi{}_\theta \wedge \sigma^\theta = d\sigma^\phi = -r\,\sin\theta\,d\theta \wedge d\phi, \tag{18.85}$$

from which it follows that

$$\omega^\theta{}_\phi = \sin\theta\,d\phi \tag{18.86}$$

and therefore that

$$d\omega^\theta{}_\phi = \cos\theta\,d\theta \wedge d\phi = \frac{\cos\theta}{r(R + r\,\cos\theta)}\,\omega \tag{18.87}$$

since $\omega^\phi{}_\theta = -\omega^\theta{}_\phi$, and where $\omega = \sigma^\theta \wedge \sigma^\phi$ is a choice of orientation. Thus, the Gaussian curvature of the torus is given by

$$K = \frac{\cos\theta}{r(R + r\,\cos\theta)}. \tag{18.88}$$

This result makes sense: The curvature in the θ direction is clearly just $1/r$, and the curvature in the ϕ direction should go like the inverse radius, or $1/(R + r\,\cos\theta)$. But why is there a factor of $\cos\theta$?

Recall that the principal curvatures tell you how the normal vector is changing. Read this sentence carefully: The principal curvatures tell you how the normal vector *to the surface* is changing along (principal) directions *in the surface*. When computing the curvature of a circle, the relevant normal vector is the outward-pointing radial vector, which lies in the plane of the circle. However, when computing the curvature of a surface *along* a circle, the relevant normal vector is that of the *surface*. For circles on the torus in the θ direction, there is no difference; these two normal vectors are the same. However, for circles in the ϕ direction, this is not the case.

Consider the circles on the top or bottom of the torus. The *surface* normal vector is always straight up (or down); it does *not* change in the ϕ direction. Thus, one principal curvature vanishes at the top and bottom of the torus, and hence so does the Gaussian curvature.

Furthermore, on the outer half of the torus, the change in the normal vector in the ϕ direction has a positive component in that direction,

whereas on the inner half of the torus, the opposite is true. This property is most easily visualized in the equatorial plane; since the surface normal vectors have opposite orientations on the inner and outer rings, the principal curvature in the ϕ direction has opposite signs there. The factor of $\cos\theta$ accounts precisely for the difference in angle between the normal vector to the surface and the (horizontal) normal vector to the circles of constant "latitude."[6]

Thus, the principal curvature in the ϕ direction, and hence the Gaussian curvature, must be positive on the outer half and negative on the inner half (and zero in between).

This should not be surprising. The Gauss–Bonnet Theorem tells us that

$$\int_T K\,\omega = 2\pi\chi(\mathbb{T}), \tag{18.89}$$

and it is easy to check by example that the Euler characteristic of a torus is zero, that is, $\chi(\mathbb{T}) = 0$. Thus, the *total* curvature of any torus must be zero, so that regions of positive curvature must be counterbalenced by regions of negative curvature. This is a *topological* statement; no matter how you twist a torus, its total curvature must be zero.

Can a torus be flat; that is, can its curvature be zero *everywhere*? Surprisingly, the answer is yes—and if you have ever played a video game, it is likely that you have already encountered a flat torus. Take a rectangular piece of paper and identify opposite edges; this is just the "wraparound" feature of many video games, in which leaving the screen on the right causes you to reenter the screen on the left, etc. The curvature is clearly zero; locally, it's just a piece of paper. And the topology is indeed that of a torus, as can be shown by experimentally determining the Euler characteristic.

Can you actually make such a torus? Not in three dimensions! This torus naturally lives in *four* dimensions, as can be seen by the following parametrization in $\mathbb{R}^4$:

$$x = r\,\cos\theta, \tag{18.90}$$

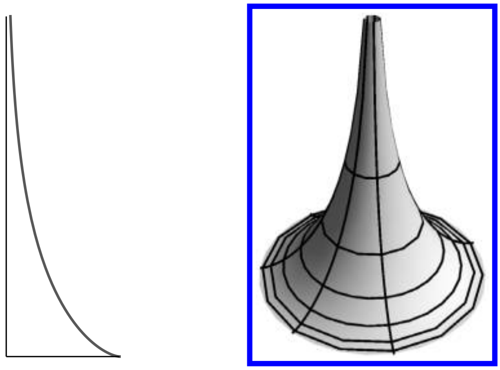
$$y = r\,\sin\theta, \tag{18.91}$$

$$z = s\,\cos\phi, \tag{18.92}$$

$$w = s\,\sin\phi, \tag{18.93}$$

which also shows why we often write the torus $\mathbb{T}$ as $\mathbb{S}^1 \times \mathbb{S}^1$, the product of two circles.

[6] A similar argument holds on a sphere, where at most latitudes the normal vector to the sphere is *not* in the plane of a circle of constant latitude.



(a) (b)

FIGURE 18.13. (a) The tractrix, whose surface of revolution is (b) the pseudosphere.

18.12 PROBLEMS

1. **Pseudosphere**

   A curve known as a *tractrix* is shown in Figure 18.13(a). As shown in Figure 18.13(b), the (upper half of the) *pseudosphere* is the surface of revolution obtained by rotating a tractrix about the vertical axis (z) and is given by

   $$x = a\sin\alpha\cos\phi, \qquad y = a\sin\alpha\sin\phi, \qquad z = -a\cos\alpha - a\ln\tan\frac{\alpha}{2},$$

   with $a =$ constant, $\alpha \in [0, \frac{\pi}{2}]$, and $\phi \in [0, 2\pi)$. The line element on the pseudosphere turns out to be

   $$ds^2 = \frac{a^2\cos^2\alpha}{\sin^2\alpha}\,d\alpha^2 + a^2\sin^2\alpha\,d\phi^2.$$

   (a) Find the (Gaussian) curvature of this surface.

   (b) Find the surface area of the (upper half of the) pseudosphere.

2. **Poincaré Half-Plane**

   The *Poincaré Half-Plane* model of hyperbolic geometry has the line element

   $$ds^2 = a^2\left(\frac{dX^2 + dY^2}{Y^2}\right),$$

   where a is a (positive) constant. Find the (Gaussian) curvature of this surface.

Chapter 19

Geodesics

19.1 Geodesics

When is a curve “straight”?

Consider the velocity vector associated with a curve, given by

$$\vec{\boldsymbol{v}} = \frac{d\vec{\boldsymbol{r}}}{dt} \tag{19.1}$$

so that

$$\vec{\boldsymbol{v}}\,dt = d\vec{\boldsymbol{r}} = \sigma^i\,\hat{\boldsymbol{e}}_i. \tag{19.2}$$

Thus,

$$\sigma^i = v^i\,dt. \tag{19.3}$$

Taking the exterior derivative of $\vec{\boldsymbol{v}}$, we obtain

$$\begin{aligned} d\vec{\boldsymbol{v}} = d(v^i\,\hat{\boldsymbol{e}}_i) &= dv^i\,\hat{\boldsymbol{e}}_i + v^i d\hat{\boldsymbol{e}}^i \\ &= (dv^j + \omega^j{}_i v^i)\,\hat{\boldsymbol{e}}_j \\ &= (dv^j + \Gamma^j{}_{ik} v^i\,\sigma^k)\,\hat{\boldsymbol{e}}_j \\ &= (dv^j + \Gamma^j{}_{ik} v^i\,v^k\,dt)\,\hat{\boldsymbol{e}}_j. \end{aligned} \tag{19.4}$$

Thus,

$$\frac{d\vec{\boldsymbol{v}}}{dt} = \left(\frac{dv^j}{dt} + \Gamma^j{}_{ik} v^i\,v^k\right)\hat{\boldsymbol{e}}_j \tag{19.5}$$

or equivalently

$$\dot{\vec{\boldsymbol{v}}} = \left(\dot{v}^j + \Gamma^j{}_{ik} v^i\,v^k\right)\hat{\boldsymbol{e}}_j. \tag{19.6}$$

Intuitively, a curve is straight if there is no acceleration, so we require

$$\dot{\vec{\boldsymbol{v}}} = 0 \tag{19.7}$$

or equivalently

$$\dot{v}^j + \Gamma^j{}_{ik} v^i\,v^k = 0. \tag{19.8}$$

A curve satisfying (19.7), and hence also (19.8), is called a *geodesic.*

19.2 Geodesics in Three Dimensions

We already have a definition of geodesics that lie in surfaces in $\mathbb{R}^3$, namely curves whose geodesic curvature vanishes. Recall that

$$\kappa_g \, ds = d\hat{\boldsymbol{T}} \cdot \hat{\boldsymbol{N}}, \tag{19.9}$$

where $\hat{\boldsymbol{T}}$ is the unit tangent vector to the curve and $\hat{\boldsymbol{N}}$ is a particular choice of normal vector to the curve. Thus, for κ_g to vanish, we must have

$$\dot{\hat{\boldsymbol{T}}} \cdot \hat{\boldsymbol{N}} = 0. \tag{19.10}$$

Since $\hat{\boldsymbol{T}}$ is a unit vector, we know that

$$\dot{\hat{\boldsymbol{T}}} \cdot \hat{\boldsymbol{T}} = 0. \tag{19.11}$$

Since we are in a two-dimensional surface, $\dot{\hat{\boldsymbol{T}}}$ has only two components. Since the $\hat{\boldsymbol{T}}$-component vanishes, the other component vanishes if and only if the vector itself is zero. Thus,

$$\kappa_g = 0 \Longleftrightarrow \dot{\hat{\boldsymbol{T}}} = 0. \tag{19.12}$$

This is almost, but not quite, our new definition of geodesic, which says that $\dot{\vec{\boldsymbol{v}}} = 0$. But

$$\vec{\boldsymbol{v}} = v \, \hat{\boldsymbol{T}}, \tag{19.13}$$

where $v = |\vec{\boldsymbol{v}}|$ is the speed, so we have

$$\dot{\vec{\boldsymbol{v}}} = \dot{v} \, \hat{\boldsymbol{T}} + v \, \dot{\hat{\boldsymbol{T}}}. \tag{19.14}$$

Since $\dot{\hat{\boldsymbol{T}}} \perp \hat{\boldsymbol{T}}$, these components are independent, and

$$\dot{\vec{\boldsymbol{v}}} = 0 \Longleftrightarrow \dot{v} = 0 \text{ and } \dot{\hat{\boldsymbol{T}}} = 0. \tag{19.15}$$

Our new definition of geodesic is therefore equivalent to the old definition, provided we traverse the curve at constant (but not necessarily unit) speed.[1]

[1] Authors differ as to whether a curve must be traversed at constant speed in order to be a geodesic. We will henceforth refer to a curve as a geodesic only if it is indeed traversed at constant speed.

19.3 EXAMPLES OF GEODESICS

THE PLANE

In rectangular coordinates, it is clear that $\dot{v}$ will be zero if and only if $\ddot{x} = 0 = \ddot{y}$; the geodesics are straight lines. Consider this same problem in polar coordinates, where

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}}. \tag{19.16}$$

Then

$$\vec{\boldsymbol{v}} = \dot{r}\,\hat{\boldsymbol{r}} + r\,\dot{\phi}\,\hat{\boldsymbol{\phi}} \tag{19.17}$$

so that

$$\begin{aligned} d\vec{\boldsymbol{v}} &= d\dot{r}\,\hat{\boldsymbol{r}} + \dot{r}\,d\hat{\boldsymbol{r}} + d(r\dot{\phi})\,\hat{\boldsymbol{\phi}} + r\dot{\phi}\,d\hat{\boldsymbol{\phi}} \\ &= d\dot{r}\,\hat{\boldsymbol{r}} + \dot{r}\,(d\phi\,\hat{\boldsymbol{\phi}}) + d(r\dot{\phi})\,\hat{\boldsymbol{\phi}} - r\dot{\phi}\,(d\phi\,\hat{\boldsymbol{r}}). \end{aligned} \tag{19.18}$$

Dividing by dt now yields

$$\begin{aligned} \dot{\vec{\boldsymbol{v}}} &= (\ddot{r} - r\dot{\phi}^2)\,\hat{\boldsymbol{r}} + \left((r\dot{\phi})^{\cdot} + \dot{r}\dot{\phi}\right)\hat{\boldsymbol{\phi}} \\ &= (\ddot{r} - r\dot{\phi}^2)\,\hat{\boldsymbol{r}} + (r\ddot{\phi} + 2\dot{r}\dot{\phi})\,\hat{\boldsymbol{\phi}} \end{aligned} \tag{19.19}$$

so that the geodesic equation in polar coordinates reduces to

$$\ddot{r} - r\dot{\phi}^2 = 0 = r\ddot{\phi} + 2\dot{r}\dot{\phi}. \tag{19.20}$$

We will discuss solution strategies for this system of ordinary differential equations in Sections 19.4 and 19.5 and content ourselves here with checking special cases. A radial line through the origin has $\phi =$ constant, so that (19.20) is satisfied *if* $\ddot{r} = 0$. But the speed in this case is just $v = |\vec{\boldsymbol{v}}| = |\dot{r}|$, so this condition says that radial lines are geodesics *if* they are traversed at constant speed.

A less obvious example is a line of the form $x =$ constant, so that

$$r\cos\phi = \text{constant}, \tag{19.21}$$

which implies that

$$\dot{r}\cos\phi - r\sin\phi\,\dot{\phi} = 0, \tag{19.22}$$

and further differentiation yields

$$(\ddot{r} - r\dot{\phi}^2)\cos\phi - (r\ddot{\phi} + 2\dot{r}\dot{\phi})\sin\phi = 0. \tag{19.23}$$

If we also assume that the line is traversed at constant speed, we must have

$$0 = \ddot{y} = (\ddot{r} - r\dot{\phi}^2)\sin\phi + (r\ddot{\phi} + 2\dot{r}\dot{\phi})\cos\phi. \tag{19.24}$$

Comparing (19.23) and (19.24), we see that (19.20) is indeed satisfied.

THE SPHERE

On the sphere, we have

$$d\vec{\boldsymbol{r}} = r\,d\theta\,\hat{\boldsymbol{\theta}} + r\,\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}} \tag{19.25}$$

so that

$$\vec{\boldsymbol{v}} = r\,\dot{\theta}\,\hat{\boldsymbol{\theta}} + r\,\sin\theta\,\dot{\phi}\,\hat{\boldsymbol{\phi}}, \tag{19.26}$$

which implies that

$$\begin{aligned} d\vec{\boldsymbol{v}} &= r\,d\dot{\theta}\,\hat{\boldsymbol{\theta}} + r\,d(\sin\theta\,\dot{\phi})\,\hat{\boldsymbol{\phi}} + r\,\dot{\theta}\,d\hat{\boldsymbol{\theta}} + r\,\sin\theta\,\dot{\phi}\,d\hat{\boldsymbol{\phi}} \\ &= r\,d\dot{\theta}\,\hat{\boldsymbol{\theta}} + r\,d(\sin\theta\,\dot{\phi})\,\hat{\boldsymbol{\phi}} + r\,\dot{\theta}\,\cos\theta\,d\phi\,\hat{\boldsymbol{\phi}} - r\,\sin\theta\,\dot{\phi}\,\cos\theta\,d\phi\,\hat{\boldsymbol{\theta}}. \end{aligned} \tag{19.27}$$

Dividing by dt yields

$$\begin{aligned} d\vec{\boldsymbol{v}} &= r(\ddot{\theta} - \sin\theta\cos\theta\,\dot{\phi}^2)\,\hat{\boldsymbol{\theta}} + r\big((\sin\theta\,\dot{\phi})^{\boldsymbol{\cdot}} + \cos\theta\,\dot{\theta}\,\dot{\phi}\big)\,\hat{\boldsymbol{\phi}} \\ &= r\,(\ddot{\theta} - \sin\theta\cos\theta\,\dot{\phi}^2)\,\hat{\boldsymbol{\theta}} + r\sin\theta\,(\ddot{\phi} + 2\cot\theta\,\dot{\theta}\,\dot{\phi})\,\hat{\boldsymbol{\phi}} \end{aligned} \tag{19.28}$$

so that the geodesic equation on the sphere reduces to

$$\ddot{\theta} - \sin\theta\cos\theta\,\dot{\phi}^2 = 0 = \ddot{\phi} + 2\cot\theta\,\dot{\theta}\,\dot{\phi}. \tag{19.29}$$

We defer a discussion of how to solve (19.29) to Sections 19.4 and 19.6 and content ourselves here with a simple example. Along the equator, we have $\theta = \pi/2$ and $\ddot{\phi} = 0$, which shows that the equator is a geodesic and hence that all great circles are geodesics.

19.4 SOLVING THE GEODESIC EQUATION

Recall the geodesic equation in polar coordinates, namely

$$\ddot{r} - r\dot{\phi}^2 = 0, \tag{19.30}$$

$$r\ddot{\phi} + 2\dot{r}\dot{\phi} = 0. \tag{19.31}$$

Divide the second equation by $r\dot{\phi}$ to obtain

$$\frac{\ddot{\phi}}{\dot{\phi}} + \frac{2\dot{r}}{r} = 0, \tag{19.32}$$

which can be integrated to yield

$$\dot{\phi} = \frac{A}{r^2}, \tag{19.33}$$

where A is a constant. Inserting (19.33) into (19.30) and multiplying by $\dot{r}$ now yields

$$\dot{r}\ddot{r} - \frac{A^2\dot{r}}{r^3} = 0, \tag{19.34}$$

which can be integrated to yield

$$\text{constant} = \dot{r}^2 + \frac{A^2}{r^2} = \dot{r}^2 + r^2\dot{\phi}^2 = |\vec{\boldsymbol{v}}|^2. \tag{19.35}$$

Thus, geodesics must be traversed at constant speed and can be classified by the constant A, which corresponds to angular momentum about the origin. It remains, of course, to solve (19.33), which we do in Section 19.5.

Solving the geodesic equation on the sphere is similar. We now start with

$$\ddot{\theta} - \sin\theta\cos\theta\,\dot{\phi}^2 = 0, \tag{19.36}$$

$$\ddot{\phi} + 2\cot\theta\,\dot{\theta}\,\dot{\phi} = 0. \tag{19.37}$$

Divide the second equation by $\dot{\phi}$ to obtain

$$\frac{\ddot{\phi}}{\dot{\phi}} + 2\cot\theta\,\dot{\theta} = 0, \tag{19.38}$$

which can be integrated to yield

$$\dot{\phi} = \frac{B}{r^2\sin^2\theta}, \tag{19.39}$$

where B is a constant (and the constant factor of r^2 is added for later convenience). Inserting this in the first equation and multiplying by $r^2\dot{\theta}$ yields

$$r^2\ddot{\theta}\,\dot{\theta} - \frac{B^2\cos\theta}{r^2\sin^3\theta}\,\dot{\theta} = 0, \tag{19.40}$$

which can be integrated to yield

$$\text{constant} = r^2\dot{\theta}^2 + \frac{B^2}{r^2\sin^2\theta} = r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2 = |\vec{\boldsymbol{v}}|^2, \tag{19.41}$$

and again we see that geodesics must be traversed at constant speed. Thus, geodesics on the sphere can be classified by the constant B, which again corresponds to (the z-component of) angular momentum. As before, it remains to solve (19.39), which we do in Section 19.6.

19.5 GEODESICS IN POLAR COORDINATES

The geodesic equation in polar coordinates reduces to the coupled differential equations

$$\dot{\phi} = \frac{\ell}{r^2}, \tag{19.42}$$

$$\dot{r}^2 = 1 - \frac{\ell^2}{r^2}, \tag{19.43}$$

where the constant ℓ is the *angular momentum* about the origin (per unit mass) and where we have used arclength (here labeled λ) as the parameter (so that the curve is traversed at unit speed). We now proceed to solve these equations.

Separating variables, we have

$$\frac{dr}{\sqrt{1 - \frac{\ell^2}{r^2}}} = \pm d\lambda \tag{19.44}$$

so that

$$\pm\lambda = \int \frac{r\,dr}{\sqrt{r^2 - \ell^2}} = \sqrt{r^2 - \ell^2} + \text{constant}, \tag{19.45}$$

and we can ignore the integration constant, as it merely shifts the origin of λ. Thus,

$$r^2 = \lambda^2 + \ell^2, \tag{19.46}$$

and we have

$$\dot{\phi} = \frac{\ell}{r^2} = \frac{\ell}{\lambda^2 + \ell^2} \tag{19.47}$$

so that

$$\phi = \int \frac{\ell\,d\lambda}{\lambda^2 + \ell^2} = \arctan\left(\frac{\lambda}{\ell}\right) + \alpha, \tag{19.48}$$

where α is an integration constant, and we conclude that

$$\tan(\phi - \alpha) = \frac{\lambda}{\ell}. \tag{19.49}$$

We can combine (19.46) and (19.49) to express r in terms of ϕ, yielding

$$\begin{aligned} r^2 &= \ell^2 \left(\frac{\lambda^2}{\ell^2} + 1\right) \\ &= \ell^2 \left(\tan^2(\phi - \alpha) + 1\right) \\ &= \ell^2 \sec^2(\phi - \alpha), \end{aligned} \tag{19.50}$$

or in other words

$$r^2 \cos^2(\phi - \alpha) = \ell^2 \tag{19.51}$$

so that

$$r \cos(\phi - \alpha) = \pm\ell, \tag{19.52}$$

where both signs are necessary to preserve the interpretation of ℓ as angular momentum (since each curve can be traversed in either direction).

Finally, using the addition formula for cosine results in

$$r \cos\phi \cos\alpha + r \sin\phi \sin\alpha = \pm\ell, \tag{19.53}$$

which is, not surprisingly, the general form of the equation of a straight line, since $x = r\cos\phi$ and $y = r\sin\phi$.

19.6 GEODESICS ON THE SPHERE

The geodesic equation on the sphere reduces to the coupled differential equations

$$\dot{\phi} = \frac{\ell}{r^2 \sin^2\theta}, \tag{19.54}$$

$$r^2\dot{\theta}^2 = 1 - \frac{\ell^2}{r^2 \sin^2\theta}, \tag{19.55}$$

where the constant ℓ is the *angular momentum* about the z-axis (per unit mass) and where we have used arclength (here labeled λ) as the parameter (so that the curve is traversed at unit speed). We now proceed to solve these equations.

We start with

$$r\dot{\theta} = \pm\frac{\sqrt{r^2\sin^2\theta - \ell^2}}{r\sin\theta}, \tag{19.56}$$

and then write

$$\frac{d\phi}{d\theta} = \frac{\dot{\phi}}{\dot{\theta}} = \frac{\pm\ell}{\sin\theta\sqrt{r^2\sin^2\theta - \ell^2}} \tag{19.57}$$

so that

$$\begin{aligned} d\phi &= \frac{\pm\ell\, d\theta}{\sin\theta\sqrt{r^2\sin^2\theta - \ell^2}} \\ &= \frac{\pm\ell\,\csc^2\theta\, d\theta}{\sqrt{r^2 - \ell^2\frac{\cos^2\theta+\sin^2\theta}{\sin^2\theta}}} \\ &= \frac{\pm\ell\,\csc^2\theta\, d\theta}{\sqrt{r^2 - \ell^2 - \ell^2\cot^2\theta}} \\ &= \frac{\mp\ell\, d\cot\theta}{\sqrt{(r^2-\ell^2) - \ell^2\cot^2\theta}}. \end{aligned} \tag{19.58}$$

This expression can now be integrated to yield

$$\phi = \mp\arcsin\left(\frac{\ell}{\sqrt{r^2-\ell^2}}\right) + \text{constant} \tag{19.59}$$

or equivalently

$$\sin(\phi-\alpha) = \mp\sqrt{\frac{\ell^2}{r^2-\ell^2}}\cot\theta. \tag{19.60}$$

Just as before, we can bring this expression to a more familiar form by expanding the left-hand side, resulting in

$$\sin\theta(\sin\phi\cos\alpha - \cos\phi\sin\alpha) \pm \sqrt{\frac{\ell^2}{r^2-\ell^2}}\cos\theta = 0, \tag{19.61}$$

which has the form of the equation of a plane through the origin, since $x = r\sin\theta\cos\phi$, $y = \sin\theta\sin\phi$, and $z = r\cos\theta$. Not surprisingly, geodesics on the sphere lie on such planes; they are great circles.

However, (19.60) is not a very practical description of geodesics on a sphere. Working in three dimensions, there is an easier way.

Any point (θ,ϕ) on a sphere is of course also a point (x,y,z) in three dimensions or equivalently a vector from the origin. The length of the circular arc connecting these points is just the radius of the sphere times the angle—which is easily obtained using the dot product.

This strategy can also be used to find an explicit expression for the geodesic connecting two given points on the sphere. We can easily find two *orthogonal* vectors whose endpoints are on the geodesic, either using projections (essentially Gram–Schmidt orthogonalization) or using the cross product twice (the first time to obtain a vector orthogonal to the plane of the geodesic, then again to obtain a second vector in that plane orthogonal to one of the given vectors). But any circle can be parametrized

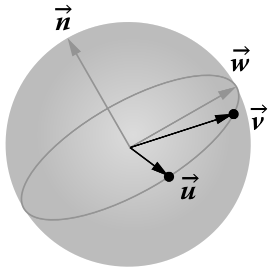


FIGURE 19.1. Using three-dimensional vectors to construct geodesics on the sphere.

in vector form as $\vec{\boldsymbol{u}}\cos\beta + \vec{\boldsymbol{w}}\sin\beta$ using any two orthogonal vectors $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{w}}$ on it. This construction is illustrated in Figure 19.1, in which the geodesic through two points $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$ on the unit sphere is constructed using $\hat{\boldsymbol{n}} = \frac{\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}}}{|\vec{\boldsymbol{u}}\times\vec{\boldsymbol{v}}|}$ and $\vec{\boldsymbol{w}} = \hat{\boldsymbol{n}} \times \vec{\boldsymbol{u}}$.

Although this construction of the geodesic equation is quite messy to use in calculations by hand, it lends itself well to programmed numerical computations.

≻ Chapter 20 ≺

Applications

In this chapter, we outline several further applications of differential forms.

20.1 The Equivalence Problem

It is easy to see that the line elements

$$ds^2 = dx^2 + dy^2, \tag{20.1}$$

$$ds^2 = dr^2 + r^2\, d\phi^2 \tag{20.2}$$

are equivalent; both represent flat Euclidean space, the first in rectangular coordinates and the second in polar coordinates. What does *equivalent* mean? That there exists a coordinate transformation taking one line element to another, in this case the standard transformation between rectangular and polar coordinates.

What about the line element

$$ds^2 = -dt^2 + t^2 \left(\frac{dr^2}{1+r^2} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2 \right) \right), \tag{20.3}$$

which was studied in Section 9.7? With the wisdom of hindsight, we can set

$$R = tr, \tag{20.4}$$

$$T = t\sqrt{1+r^2}, \tag{20.5}$$

after which the line element becomes

$$ds^2 = -dT^2 + dR^2 + R^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2 \right) \tag{20.6}$$

so that this line element is nothing more than Minkowski space in funny coordinates. But what if we didn't know about this coordinate transformation? Could we still have determined that this line element is that of Minkowski space?

The answer in this case is yes: Compute the curvature 2-forms; they vanish. There is only one flat geometry in each dimension and signature, so there must be a transformation from the original coordinates to Minkowski coordinates.

However, the general question of when two metrics are equivalent is much harder. The history of general relativity is full of rediscoveries of known solutions of Einstein's equation, but in new coordinates, where the solution was not immediately recognized. Is there an algorithm for determining equivalence? It was this question that led Christoffel to introduce connections! (See [15] for further discussion.)

In fact there is such an algorithm, at least in principle, and it goes something like this: Compute the components of the (Riemann) curvature tensor for both metrics. Now compute their (tensor, i.e., "covariant") derivatives. Repeat until you have computed all derivatives of order 20. Compare your two lists of derivatives. If they are the same, that is, if they can be made to agree after some simple algebra, then yes, the metrics are equivalent. If not, they are not. This last step is not easy, as it requires putting both lists into some sort of canonical order so that they can be compared. However, that is not the main stumbling block.

How many components does the Riemann tensor have in 4 dimensions? That is, how many independent components do the curvature 2-forms have? The Riemann tensor is a 4-index object, so it has $4^4 = 256$ components, although symmetry reduces that number significantly. But each derivative can be in any of 4 directions, so the number of derivatives of order 20 is something like $4^{4+20} = 281{,}474{,}976{,}710{,}656$. That's 281 trillion... Again, symmetry will reduce this number significantly, and the actual number of independent components through order 20 turns out to be 79,310. But that's still too many to make the comparison realistic other than in highly symmetric situations such as flat space. What can we do?

The first simplification is to insist that all computations be done in an orthonormal frame. This turns out to reduce the order of differentiation needed from 20 to 10. That reduces the number of highest-order terms to something like $4^{4+10} = 268{,}435{,}456$, and symmetry reduces the number of independent components through order 10 to 8690.

Further simplification turns out to be possible by using spinors rather than vectors. The relationship between spinors and differential forms is discussed briefly in Section 20.3. In a certain sense, spinors are the "square roots" of vectors, and this turns out to reduce the order of differentiation to 7 and the number of independent components through order 7 to 3156.

A few thousand terms may still seem like a lot, but in practice things are even easier, as no solutions to Einstein's equation are known for which the algorithm requires more than 3 derivatives, reducing the number of independent components to a mere 430. Most of the known solutions have indeed been classified, in the sense of being used to generate a (long) list of components and derivatives in canonical form that can be compared with a candidate new solution, and much of this work has been automated using purpose-designed algebraic computing software.

Further discussion of these ideas can be found in [16].

20.2 LAGRANGIANS

A *Lagrangian density* $\mathcal{L}$ is just an integrand, that is, an n-form. The corresponding *action* $\mathcal{S}$ is just the integral of the Lagrangian density, typically over the entire space. The goal is to choose the Lagrangian involving physical fields so that its extrema correspond to physically interesting conditions on those fields.

The simplest example is the action for a scalar field Φ, which is given by

$$\mathcal{L} = d\Phi \wedge *d\Phi \tag{20.7}$$

so that the corresponding action is

$$\mathcal{S} = \int_R d\Phi \wedge *d\Phi, \tag{20.8}$$

where Φ is a function, that is, a 0-form. We *vary* this action by considering a 1-parameter family Φ_λ of fields with

$$\Phi_0 = \Phi. \tag{20.9}$$

We then set

$$\delta\Phi = \left.\frac{d\Phi_\lambda}{d\lambda}\right|_{\lambda=0} \tag{20.10}$$

so that

$$\Phi_\lambda \approx \Phi_0 + \lambda\,\delta\Phi \tag{20.11}$$

to first order. This process is often written as

$$\Phi \longmapsto \Phi + \delta\Phi, \tag{20.12}$$

where $\delta\Phi$ is assumed to be small, so that higher-order terms in $\delta\Phi$ can be ignored.

We can now compute

$$\begin{aligned}\left.\frac{d\mathcal{S}}{d\lambda}\right|_{\lambda=0} &= \int_R d(\delta\Phi)\wedge *d\Phi + d\Phi\wedge *d(\delta\Phi)\\ &= \int_R 2\,d(\delta\Phi)\wedge *d\Phi\\ &= 2\int_R \big(d(\delta\Phi *d\Phi) - \delta\Phi\, d{*}d\Phi\big)\\ &= 2\int_{\partial R}\delta\Phi *d\Phi - 2\int_R \delta\Phi\, d{*}d\Phi, \end{aligned} \tag{20.13}$$

where we have used integration by parts and Stokes' Theorem in the last two steps. We now require that the integral (20.13) vanish for *any* variation $\delta\Phi$. By first choosing the support of the variation to be away from the boundary ∂R, we see that each integrand must vanish separately. Considering the second integral first, we must have

$$d{*}d\Phi = 0 \tag{20.14}$$

on R, a statement which is often written as

$$\frac{\delta\mathcal{S}}{\delta\Phi} = -2\,d{*}d\Phi = 0. \tag{20.15}$$

Thus, our original Lagrangian leads to the requirement that our scalar field satisfy Laplace's equation $\Delta\Phi = 0$—or, in the case of Minkowski space, the wave equation $\Delta\Phi = \ddot{\Phi}$.

One often assumes that the boundary term does not contribute, either by requiring that $\Phi = 0$ there (such as by imposing suitable falloff conditions to infinity) or by specifying Φ at the boundary (so that $\delta\Phi = 0$ there). In the absence of these assumptions, we can choose the variation $\delta\Phi$ to have (small) support on ∂R, which forces the first integrand to vanish as well, so that

$$*d\Phi = 0 \tag{20.16}$$

on ∂R. But $\int *F$ is the flux of F across the surface, so (20.16) tells us that the derivative of Φ in the direction perpendicular to the boundary must vanish on the boundary.

A similar argument can be used to derive Maxwell's equations starting from the Lagrangian

$$\mathcal{L} = F\wedge *F = dA\wedge *dA, \tag{20.17}$$

where F is the electromagnetic 2-form and A is the 4-potential, and Einstein's equation starting from the Lagrangian

$$\mathcal{L} = \Omega_{ij} \wedge *(\sigma^i \wedge \sigma^j), \tag{20.18}$$

where the Ω_{ij} are, of course, the curvature 2-forms (with an index lowered).

20.3 SPINORS

We now allow ourselves to formally add differential forms of different ranks and introduce a new product on the resulting space. For any function f and 1-forms α, β, define

$$f \vee \alpha = f\alpha, \tag{20.19}$$

$$\alpha \vee \beta = \alpha \wedge \beta + g(\alpha, \beta). \tag{20.20}$$

The operation $\vee$ is called the *Clifford product* and is read as "vee." We extend the Clifford product to (formal sums of) differential forms of all ranks by requiring associativity. It is then a straightforward exercise to work out that, for example,

$$\alpha \vee \beta \vee \gamma = \alpha \wedge \beta \wedge \gamma + g(\alpha, \beta)\gamma - g(\alpha, \gamma)\beta + g(\beta, \gamma)\alpha \tag{20.21}$$

for any 1-forms α, β, γ. The space of formal sums of differential forms under the Clifford product is a *Clifford algebra*.

It turns out that the algebra we have just constructed is precisely the algebra of gamma matrices used in quantum field theory! All one needs to do is make the association

$$\gamma^i \longleftrightarrow \sigma^i \tag{20.22}$$

and then identify matrix multiplication of gamma matrices with the Clifford product of differential forms. So our orthonormal basis $\{\sigma^i\}$ plays the role of the gamma matrices, and this construction is particularly simple in rectangular coordinates, where $\sigma^i = dx^i$. Spinors can be thought of as the objects acted upon by gamma matrices, which in turn can be reinterpreted in terms of the gamma matrices themselves. That's not quite the whole story, as spinors actually live in certain subspaces of Clifford algebras. Nonetheless, the study of spinors is therefore the study of Clifford algebras, for which we can use differential forms.

An example of the power of this approach is obtained by considering the *Kähler operator* $d - \delta$, where

$$\delta = (-1)^p *^{-1} d*, \tag{20.23}$$

where $*^{-1}$ is of course the same as $*$ up to sign (and where this δ has nothing to do with the variational δ introduced in Section 20.2). Since $\delta^2 = 0$, we also have

$$(d - \delta)^2 = -d\delta - \delta d. \tag{20.24}$$

Applying this operator to a function f yields

$$(d - \delta)^2 f = -d\delta f - \delta df = 0 + *^{-1} d{*}df = \Delta f \tag{20.25}$$

so that the Kähler operator is a sort of square root of the Laplacian—or, in Minkowski space, the square root of the wave equation. But that is precisely the role of the Dirac equation, the fundamental equation for spinor fields: It serves as a square root of the wave equation, also known as the (massless) Klein–Gordon equation. The Dirac equation, which describes both electrons and quarks, can therefore be rewritten in terms of differential forms. A detailed treatment of spinors in this language can be found in [17].

20.4 TOPOLOGY

A p-form α is called *closed* if its exterior derivative vanishes, and *exact* if it is the derivative of a $(p-1)$-form, that is,

$$\alpha \text{ closed} \iff d\alpha = 0, \tag{20.26}$$

$$\alpha \text{ exact} \iff \exists\beta : \alpha = d\beta. \tag{20.27}$$

Equivalently, α is closed if it lies in the *kernel* $\mathrm{Ker}(d)$ of d, and is exact if it lies in the *image* $\mathrm{Im}(d)$ of d.

Exact differentials such as $x\,dx + y\,dy = d(x^2 + y^2)$ are also exact as 1-forms and correspond (via $F = \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{r}}$) to conservative vector fields, namely vector fields that are the gradient of some function, that is, $\vec{\boldsymbol{F}} = \vec{\boldsymbol{\nabla}} f$. The *Poincaré Lemma* states that an exact differential form is also closed, which follows immediately from $d^2 = 0$, and which includes the vector calculus identities $\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{\nabla}} f = \vec{\boldsymbol{0}}$ and $\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} = 0$ as special cases.

The converse of the Poincaré Lemma[1] is *not* true in general, but requires additional assumptions about the topology of the underlying space. One version states that if R is a simply connected region of $\mathbb{R}^n$, then all closed p-forms on R are also exact. (A *simply connected* region is one whose points

[1] Rather confusingly, the converse of the Poincaré Lemma is often referred to as the Poincaré Lemma.

can be connected by curves and in which any closed curve can be shrunk to a point. Loosely speaking, a simply connected region "has no holes.")

For example, any closed form in Euclidean $\mathbb{R}^3$ is also exact, that is,

$$d\alpha = 0 \Longrightarrow \exists \beta : \alpha = d\beta \tag{20.28}$$

for any p-form α. If α is a 1-form, (20.28) is nothing more than the curl test for conservative vector fields: If $\vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{F}} = \vec{\boldsymbol{0}}$, then $\vec{\boldsymbol{F}} = \vec{\boldsymbol{\nabla}} f$ for some *potential function* f. Similarly, if α is a 2-form, then (20.28) corresponds to the statement that if $\vec{\boldsymbol{\nabla}} \cdot \vec{\boldsymbol{F}} = 0$, then $\vec{\boldsymbol{F}} = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{A}}$ for some *vector potential* $\vec{\boldsymbol{A}}$.

Not surprisingly, the $(p-1)$-form β in (20.27) is called a *potential* for the p-form α. Potentials are not unique, since we can always add an exact $(p-1)$-form to β, that is,

$$\alpha = d\beta \Longrightarrow \alpha = d(\beta + d\gamma), \tag{20.29}$$

a special case of which is the addition of an integration constant to an indefinite integral, corresponding to

$$d(f + \text{constant}) = df. \tag{20.30}$$

So the *failure* of closed forms to be exact is an indication of nontrivial topology!

For example, consider the unit circle, $\mathbb{S}^1$. The 1-form $d\phi$ is clearly closed, since there are no nonzero 2-forms on $\mathbb{S}^1$. We have

$$\int_{\mathbb{S}^1} d\phi = 2\pi \neq 0, \tag{20.31}$$

which, as discussed in Section 16.7, tells us that $d\phi$ cannot be exact! But surely $d\phi$ is $d(\phi)$? This assertion is in fact false, and for a simple reason: There is no smooth function "ϕ" *on the entire circle* whose differential is $d\phi$. Rather, it takes *two* such functions, call them ϕ_1 running from 0 to 2π and ϕ_2 running from $-\pi$ to π *with the endpoints excluded.* Somewhat remarkably, $d\phi$ *is* well-defined everywhere, since $d\phi_1 = d\phi_2$ wherever both are defined. The problem is that ϕ is not well-defined. So $d\phi$ is not actually d of any (single) function.

These ideas can be generalized into a characterization of topological structure known as *de Rham cohomology.* Any 1-form on $\mathbb{S}^1$ takes the form

$$\alpha = f\, d\phi \tag{20.32}$$

so that

$$\int_{\mathbb{S}^1} \alpha = \int_0^{2\pi} f\, d\phi = c_\alpha, \tag{20.33}$$

where the last equality defines the constant c_α. But then

$$\gamma = \alpha - \frac{c_\alpha}{2\pi} \tag{20.34}$$

does integrate to zero and therefore can be (and in fact is) exact; that is, $\gamma = dg$, so that

$$\alpha = dg + \frac{c_\alpha}{2\pi} \tag{20.35}$$

for some function g. We can therefore divide up the 1-forms on $\mathbb{S}^1$ into equivalence classes based on the value of c_α by considering two 1-forms to be equivalent if they differ by an exact differential such as dg. This amounts to considering the quotient space

$$\mathrm{Ker}(d)/\mathrm{Im}(d) = \{\text{closed forms}\}/\{\text{exact forms}\}, \tag{20.36}$$

which is called the (first) *de Rham cohomology class*. The fact that this space is one-dimensional in this case, consisting of all real numbers c_α, turns out to imply that there is one "hole," a statement about the topology of the circle.

20.5 INTEGRATION ON THE SPHERE

Start by working directly in $\mathbb{R}^3$ and using rectangular coordinates. Choose any 1-form β in $\mathbb{R}^3$, expressed in terms of rectangular coordinates. Compute $\alpha = d\beta$. If you get zero, start over; β should *not* be exact. Now switch to spherical coordinates and evaluate α on the unit sphere, $\mathbb{S}^2$. Use what you know!

Now evaluate β on the unit sphere, then take d of the result. Did you get the same answer? That is, does exterior differentiation commute with the process of evaluating a differential form on a surface?

The answer should be yes. The result of "evaluating" a differential form on a surface is called the *pullback* of the form from the larger space to the surface, and pullbacks do commute with d.

So what happens if you integrate α over $\mathbb{S}^2$? To integrate a differential form, simply drop the wedge products and do the integral—and check whether the orientation agrees with the one you want, inserting a minus sign if it doesn't. Since $\alpha = d\beta$ *on the sphere*, you can use Stokes' Theorem

as discussed in Section 16.7; your integral must therefore evaluate to zero. Did it?

Let's try this again, working directly on $\mathbb{S}^2$, using the standard orientation $\omega = \sin\theta\, d\theta \wedge d\phi$. We know that

$$\int_{\mathbb{S}^2} \omega = 4\pi \neq 0 \tag{20.37}$$

so that, just as in Section 20.4, ω cannot be exact. Yet clearly,

$$\omega = d(-\cos\theta\, d\phi). \tag{20.38}$$

What is going on here?

The function $\cos\theta$ is well-defined on $\mathbb{S}^2$, so the problem must be with $d\phi$. Intuitively, we expect $\sigma^\phi = \sin\theta\, d\phi$ to be well-defined; after all, it is a normalized basis 1-form. But

$$\cos\theta\, d\phi = \cot\theta\, \sigma^\phi, \tag{20.39}$$

and $\cot\theta$ is infinite at the poles. Put differently, $d\phi$ is not defined at the poles, although the combination $\sin\theta\, d\phi$ is reasonably behaved.

Another way to see this is to go back to $\mathbb{R}^3$ and use rectangular coordinates. We have

$$x\, dy - y\, dx = r^2\, \sin^2\theta\, d\phi \tag{20.40}$$

or equivalently

$$r\, \sin\theta\, d\phi = \cos\phi\, dy - \sin\phi\, dx, \tag{20.41}$$

which shows explicitly that $r\, \sin\theta\, d\phi$ has magnitude 1 everywhere, including when $x = 0 = y$ (even though the limit does depend on ϕ). Multiplying these expressions by $\cot\theta$ clearly leads to a 1-form that is not well behaved, either in $\mathbb{R}^3$ or on $\mathbb{S}^2$, namely $r\, \cos\theta\, d\phi$.

≻ Appendix A ≺

Detailed Calculations

A.1 Coordinate Symmetries

We show here that coordinate directions in which the metric (line element) doesn't change always correspond to Killing vectors.

Theorem: *Suppose that* $\{x = y^0, y^1, ...\}$ *are orthogonal coordinates, so that*[1]

$$d\vec{\boldsymbol{r}} = h\,dx\,\hat{\boldsymbol{x}} + \sum_i h_i\,dy^i\,\hat{\boldsymbol{y}}^i, \tag{A.1}$$

and suppose further that the coefficients $h = h_0, h_1, ...$ *do not depend on* x*; that is, suppose that*

$$\frac{\partial h}{\partial x} = 0 = \frac{\partial h_i}{\partial x}. \tag{A.2}$$

Then $\vec{\boldsymbol{X}} = h\,\hat{\boldsymbol{x}}$ *is a Killing vector.*

Proof:[2] Our orthonormal basis of 1-forms is $\{\sigma^x = h\,dx, \sigma^i = h_i\,dy^i\}$. The structure equation for $d\sigma^x$ is

$$\begin{aligned} 0 &= d\sigma^x + \sum_i \omega^x{}_i \wedge \sigma^i \\ &= dh \wedge dx + \sum_i \omega^x{}_i \wedge h_i\,dy^i. \end{aligned} \tag{A.3}$$

Writing

$$\omega^x{}_i = \Gamma^x{}_{ix}\,\sigma^x + \sum_j \Gamma^x{}_{ij}\,\sigma^j \tag{A.4}$$

and collecting terms *not* involving dx, we obtain

$$\sum_{i,j} \Gamma^x{}_{ij}\,\sigma^j \wedge \sigma^i = 0. \tag{A.5}$$

[1] Throughout this section, all sums are shown explicitly; repeated indices are *not* summed over otherwise. Furthermore, x is *not* included in the sums, that is, $i \neq 0$.

[2] This proof is easier using coordinate bases rather than the orthonormal bases used here, but that is beyond the scope of this book.

On the other hand, using the results (and notation) of Section 17.8, we have

$$\begin{aligned}2\Gamma_{xij} &= g(d\sigma_x, \sigma_i \wedge \sigma_j) - g(d\sigma_j, \sigma_x \wedge \sigma_i) + g(d\sigma_i, \sigma_j \wedge \sigma_x)\\ &= g(d\sigma_x, \sigma_i \wedge \sigma_j), \end{aligned} \tag{A.6}$$

where

$$\Gamma_{xij} = \epsilon\,\Gamma^x{}_{ij} \tag{A.7}$$

and

$$\epsilon = \hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} = \pm 1. \tag{A.8}$$

(The last two terms in the first equality in (A.6) vanish, because $d\sigma_i = dh_i \wedge dy^i$ contains no dx terms by assumption and is hence orthogonal to any 2-form containing σ^x.) Since the final expression in (A.6) is antisymmetric in the indices i and j, so is $\Gamma^x{}_{ij}$; that is,

$$\Gamma^x{}_{ji} = -\Gamma^x{}_{ij}. \tag{A.9}$$

But (A.5) implies that $\Gamma^x{}_{ij}$ is *symmetric* in i and j, and must therefore vanish.

Thus, $\omega^x{}_i$ is proportional to $\sigma^x = h\,dx$, and we can therefore conclude that

$$\omega^x{}_i = \frac{1}{h_i}\frac{\partial h}{\partial y^i}\,dx. \tag{A.10}$$

Since

$$d\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}} = \frac{1}{2} d(\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{x}}) = 0, \tag{A.11}$$

we have

$$\begin{aligned} d\hat{\boldsymbol{x}} \cdot d\vec{\boldsymbol{r}} &= 0 + \sum_i h_i\,dy^i\,d\hat{\boldsymbol{x}} \cdot \hat{\boldsymbol{y}}^i = \sum_i h_i\,dy^i\,\omega_{ix}\\ &= -\epsilon \sum_i h_i\,dy^i\,\omega^x{}_i = -\epsilon\,dh\,dx \end{aligned} \tag{A.12}$$

and of course

$$\hat{\boldsymbol{x}} \cdot d\vec{\boldsymbol{r}} = \epsilon\,h\,dx. \tag{A.13}$$

Putting this all together, we have

$$\begin{aligned} d\vec{\boldsymbol{X}} \cdot d\vec{\boldsymbol{r}} &= d(h\,\hat{\boldsymbol{x}}) \cdot d\vec{\boldsymbol{r}}\\ &= dh\,\hat{\boldsymbol{x}} \cdot d\vec{\boldsymbol{r}} + h\,d\hat{\boldsymbol{x}} \cdot d\vec{\boldsymbol{r}}\\ &= \epsilon\,h\,dh\,dx - \epsilon\,h\,dh\,dx\\ &= 0. \end{aligned} \tag{A.14}$$

Thus, $\vec{\boldsymbol{X}}$ is indeed a Killing vector, as claimed. □

A.2 GEODESIC DEVIATION: DETAILS

As in Section 7.3, let $\vec{\boldsymbol{v}} = \dot{\vec{\boldsymbol{r}}}$ denote the velocity vectors of a family of geodesics, and $\vec{\boldsymbol{u}} = \vec{\boldsymbol{r}}'$ the separation vector between them.[3] Then along the geodesics we have

$$\vec{\boldsymbol{v}}\,d\tau = d\vec{\boldsymbol{r}}, \tag{A.15}$$

$$v^k\,d\tau = \sigma^k, \tag{A.16}$$

and similar expressions hold along curves with tangent vector $\vec{\boldsymbol{u}}$, namely

$$\vec{\boldsymbol{u}}\,ds = d\vec{\boldsymbol{r}}, \tag{A.17}$$

$$u^k\,ds = \sigma^k. \tag{A.18}$$

Let $\vec{\boldsymbol{w}} = w^i\hat{\boldsymbol{e}}_i$ be a vector field. Then as usual we have

$$d\vec{\boldsymbol{w}} = \left(dw^i + \omega^i{}_j\,w^j\right)\,\hat{\boldsymbol{e}}_i = \left(dw^i + \Gamma^i{}_{jk}w^j\sigma^k\right)\,\hat{\boldsymbol{e}}_i, \tag{A.19}$$

where the $\Gamma^i{}_{jk}$ are the components of the connection 1-forms $\omega^i{}_j$. We can now compute

$$\dot{\vec{\boldsymbol{w}}} = \left(\dot{w}^k + \Gamma^k{}_{ij}w^iv^j\right)\,\hat{\boldsymbol{e}}_k, \tag{A.20}$$

$$\vec{\boldsymbol{w}}' = \left((w^k)' + \Gamma^k{}_{ij}w^iu^j\right)\,\hat{\boldsymbol{e}}_k. \tag{A.21}$$

Since $\vec{\boldsymbol{v}}$ corresponds to a geodesic, we have

$$\dot{v}^k + \Gamma^k{}_{ij}v^iv^j = 0, \tag{A.22}$$

and by assumption we have

$$\dot{\vec{\boldsymbol{u}}} = \vec{\boldsymbol{v}}' \tag{A.23}$$

so that

$$\dot{u}^k - (v^k)' = \Gamma^k{}_{ij}(v^iu^j - u^iv^j). \tag{A.24}$$

We now calculate $\ddot{\vec{\boldsymbol{u}}} = (\vec{\boldsymbol{v}}')^{\cdot}$. Using (A.20) and (A.21), we have

$$\vec{\boldsymbol{v}}' = \left((v^k)' + \Gamma^k{}_{ij}v^iu^j\right)\,\hat{\boldsymbol{e}}_k \tag{A.25}$$

and then

$$\begin{aligned}(\vec{\boldsymbol{v}}')^{\cdot} &= \left[\left((v^k)' + \Gamma^k{}_{ij}v^iu^j\right)^{\cdot} + \Gamma^k{}_{ij}\left((v^i)' + \Gamma^i{}_{mn}v^mu^n\right)v^j\right]\,\hat{\boldsymbol{e}}_k \\ &= \Big[(\dot{v}^k)' + \dot{\Gamma}^k{}_{ij}v^iu^j + \Gamma^k{}_{ij}\left(\dot{v}^iu^j + v^i\dot{u}^j\right) \\ &\qquad + \Gamma^k{}_{ij}v^j\left((v^i)' + \Gamma^i{}_{mn}v^mu^n\right)\Big]\,\hat{\boldsymbol{e}}_k.\end{aligned} \tag{A.26}$$

[3] We usually think of $\vec{\boldsymbol{v}}$ as corresponding to a *timelike* geodesic, and $\vec{\boldsymbol{u}}$ to an orthogonal, spacelike direction, although this is not necessary. We nonetheless use τ as the parameter along the geodesics, and s as the parameter that labels the geodesics.

Inserting (A.22) and (A.24) leads to

$$
\begin{aligned}
(\vec{\boldsymbol{v}}')^{\cdot} &= \Big[-(\Gamma^k{}_{ij})' v^i v^j - \Gamma^k{}_{ij}\left((v^i)'v^j + v^i(v^j)'\right) + \dot{\Gamma}^k{}_{ij} v^i u^j \\
&\quad + \Gamma^k{}_{ij}\left(\dot{v}^i u^j + v^i \dot{u}^j\right) + \Gamma^k{}_{ij} v^j \left((v^i)' + \Gamma^i{}_{mn} v^m u^n\right)\Big]\,\hat{\boldsymbol{e}}_k \\
&= \Big[-(\Gamma^k{}_{ij})' v^i v^j + \dot{\Gamma}^k{}_{ij} v^i u^j + \Gamma^k{}_{ij}\Gamma^j{}_{mn} v^i (v^m u^n - u^m v^n) \\
&\quad + \Gamma^k{}_{ij}\Gamma^i{}_{mn} v^m (v^j u^n - u^j v^n)\Big]\,\hat{\boldsymbol{e}}_k \\
&= \Big[-(\Gamma^k{}_{ij})' v^i v^j + \dot{\Gamma}^k{}_{ij} v^i u^j \\
&\quad + \left(\Gamma^k{}_{ij}\Gamma^j{}_{mn} + \Gamma^k{}_{jm}\Gamma^j{}_{in}\right) v^i (v^m u^n - u^m v^n)\Big]\,\hat{\boldsymbol{e}}_k. \qquad \text{(A.27)}
\end{aligned}
$$

The first thing to note about the factor in square brackets is that it depends only on $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$, not on their derivatives. To recognize this factor, we begin by computing the curvature 2-forms.

Taking the exterior derivative of

$$\omega^i{}_j = \Gamma^i{}_{jk}\,\sigma^k, \qquad \text{(A.28)}$$

we obtain

$$
\begin{aligned}
\Omega^i{}_j &= d\omega^k{}_i + \omega^k{}_j \wedge \omega^j{}_i \\
&= d\Gamma^k{}_{ij} \wedge \sigma^j + \Gamma^k{}_{ij}\, d\sigma^j + \omega^k{}_j \wedge \omega^j{}_i \\
&= d\Gamma^k{}_{ij} \wedge \sigma^j - \Gamma^k{}_{ij}\, \omega^j{}_m \wedge \sigma^m + \omega^k{}_j \wedge \omega^j{}_i \\
&= d\Gamma^k{}_{ij} \wedge \sigma^j - \Gamma^k{}_{ij}\Gamma^j{}_{mn}\, \sigma^n \wedge \sigma^m + \Gamma^k{}_{jm}\Gamma^j{}_{in}\, \sigma^m \wedge \sigma^n. \qquad \text{(A.29)}
\end{aligned}
$$

We are interested in the coefficient of $ds \wedge d\tau$, so we can assume without loss of generality that only τ and s are changing, so that

$$df = \dot{f}\, d\tau + f'\, ds \qquad \text{(A.30)}$$

for any function f, and any 2-form must be proportional to $ds \wedge d\tau$. Thus,

$$
\begin{aligned}
\Omega^k{}_i &= \left(\dot{\Gamma}^k{}_{ij} d\tau + (\Gamma^k{}_{ij})' ds\right) \wedge \sigma^j + \left(\Gamma^k{}_{jm}\Gamma^j{}_{in} + \Gamma^k{}_{ij}\Gamma^j{}_{mn}\right) \sigma^m \wedge \sigma^n \\
&= \Big[-\dot{\Gamma}^k{}_{ij} u^j + (\Gamma^k{}_{ij})' v^j + \left(\Gamma^k{}_{jm}\Gamma^j{}_{in} + \Gamma^k{}_{ij}\Gamma^j{}_{mn}\right)(u^m v^n - v^m u^n)\Big] \\
&\qquad ds \wedge d\tau \\
&= \frac{1}{2} R^k{}_{imn}\, \sigma^m \wedge \sigma^n = \frac{1}{2} R^k{}_{imn}\,(u^m v^n - v^m u^n)\, ds \wedge d\tau \\
&= R^k{}_{imn}\, u^m v^n\, ds \wedge d\tau. \qquad \text{(A.31)}
\end{aligned}
$$

Comparing (A.23), (A.27), and (A.31) we finally conclude that

$$\ddot{\vec{\boldsymbol{u}}} = -R^k{}_{imn} v^i u^m v^n \hat{\boldsymbol{e}}_k, \tag{A.32}$$

as claimed. This expression for the relative acceleration of nearby geodesics ($\vec{\boldsymbol{v}}$) separated in the $\vec{\boldsymbol{u}}$ direction is called the equation of *geodesic deviation.*

A.3 SCHWARZSCHILD CURVATURE

The Schwarzschild line element is

$$ds^2 = -f\,dt^2 + \frac{dr^2}{f} + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2, \tag{A.33}$$

where we have written f for the function

$$f(r) = 1 - \frac{2m}{r}, \tag{A.34}$$

with derivatives

$$f' = \frac{df}{dr} = \frac{2m}{r^2}, \tag{A.35}$$

$$f'' = -\frac{4m}{r^3}. \tag{A.36}$$

It now follows immediately that

$$d\vec{\boldsymbol{r}} = \sqrt{f}\,dt\,\hat{\boldsymbol{t}} + \frac{dr}{\sqrt{f}}\,\hat{\boldsymbol{r}} + r\,d\theta\,\hat{\boldsymbol{\theta}} + r\,\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}, \tag{A.37}$$

and the basis 1-forms are

$$\sigma^t = \sqrt{f}\,dt, \tag{A.38}$$

$$\sigma^r = \frac{dr}{\sqrt{f}}, \tag{A.39}$$

$$\sigma^\theta = r\,d\theta, \tag{A.40}$$

$$\sigma^\phi = r\,\sin\theta\,d\phi. \tag{A.41}$$

The structure equations therefore become

$$d\sigma^t = \frac{1}{2}\frac{f'}{\sqrt{f}}\,dr\wedge dt = -\omega^t{}_m \wedge \sigma^m, \tag{A.42}$$

$$d\sigma^r = 0 = -\omega^r{}_m \wedge \sigma^m, \tag{A.43}$$

$$d\sigma^\theta = dr\wedge d\theta = -\omega^\theta{}_m \wedge \sigma^m, \tag{A.44}$$

$$d\sigma^\phi = \sin\theta\,dr\wedge d\phi + r\,\cos\theta\,d\theta\wedge d\phi = -\omega^\phi{}_m \wedge \sigma^m. \tag{A.45}$$

These equations suggest that

$$\omega^t{}_r = \frac{1}{2} f'\, dt = \frac{1}{2} \frac{f'}{\sqrt{f}}\, \sigma^t, \tag{A.46}$$

$$\omega^\theta{}_r = \sqrt{f}\, d\theta = \frac{\sqrt{f}}{r}\, \sigma^\theta, \tag{A.47}$$

$$\omega^\phi{}_r = \sqrt{f}\, \sin\theta\, d\phi = \frac{\sqrt{f}}{r}\, \sigma^\phi, \tag{A.48}$$

$$\omega^\phi{}_\theta = \cos\theta\, d\phi = \frac{1}{r} \cot\theta\, \sigma^\phi, \tag{A.49}$$

and it is easy to check that these educated guesses actually do satisfy the structure equations, remembering that

$$\omega^r{}_t = \omega^t{}_r, \qquad \omega^r{}_\theta = -\omega^\theta{}_r, \qquad \omega^r{}_\phi = -\omega^\phi{}_r, \qquad \omega^\theta{}_\phi = -\omega^\phi{}_\theta \tag{A.50}$$

but setting all other connection 1-forms to zero. Since we are guaranteed a unique solution, we have found our connection 1-forms.

The curvature 2-forms are now given by

$$\Omega^t{}_r = d\omega^t{}_r + \omega^t{}_m \wedge \omega^m{}_r = \frac{1}{2} f''\, dr \wedge dt = -\frac{1}{2} f''\, \sigma^t \wedge \sigma^r, \tag{A.51}$$

$$\Omega^t{}_\theta = d\omega^t{}_\theta + \omega^t{}_m \wedge \omega^m{}_\theta = -f' \sqrt{f}\, dt \wedge d\theta = -\frac{f'}{2r} \sigma^t \wedge \sigma^\theta, \tag{A.52}$$

$$\Omega^t{}_\phi = d\omega^t{}_\phi + \omega^t{}_m \wedge \omega^m{}_\phi = -f' \sqrt{f}\, \sin\theta\, dt \wedge d\phi = -\frac{f'}{2r} \sigma^t \wedge \sigma^\phi, \tag{A.53}$$

$$\Omega^\theta{}_r = d\omega^\theta{}_r + \omega^\theta{}_m \wedge \omega^m{}_r = \frac{1}{2} \frac{f'}{\sqrt{f}}\, dr \wedge d\theta = -\frac{f'}{2r}\, \sigma^\theta \wedge \sigma^r, \tag{A.54}$$

$$\begin{aligned}
\Omega^\phi{}_r &= d\omega^\phi{}_r + \omega^\phi{}_m \wedge \omega^m{}_r \\
&= \frac{1}{2} \frac{f'}{\sqrt{f}}\, \sin\theta\, dr \wedge d\phi + \sqrt{f}\, \cos\theta\, d\theta \wedge d\phi + \sqrt{f}\, \cos\theta\, d\phi \wedge d\theta \\
&= -\frac{f'}{2r}\, \sigma^\phi \wedge \sigma^r, &\text{(A.55)}\\
\Omega^\phi{}_\theta &= d\omega^\phi{}_\theta + \omega^\phi{}_m \wedge \omega^m{}_\theta \\
&= -\sin\theta\, d\theta \wedge d\phi - f \sin\theta\, d\phi \wedge d\theta = \frac{1-f}{r^2} \sigma^\phi \wedge \sigma^\theta. &\text{(A.56)}
\end{aligned}$$

Again we have

$$\Omega^r{}_t = \Omega^t{}_r, \qquad \Omega^r{}_\theta = -\Omega^\theta{}_r, \qquad \Omega^r{}_\phi = -\Omega^\phi{}_r, \qquad \Omega^\theta{}_\phi = -\Omega^\phi{}_\theta \tag{A.57}$$

as well as

$$\Omega^\theta{}_t = \Omega^t{}_\theta, \quad \Omega^\phi{}_t = \Omega^t{}_\phi. \tag{A.58}$$

Substituting for f, we obtain the Schwarzschild curvature as given in Section 7.4.

The (nonzero, independent) components of the Riemann tensor are therefore given by

$$R^t{}_{rtr} = -\frac{f''}{2}, \tag{A.59}$$

$$R^t{}_{\theta t\theta} = R^t{}_{\phi t\phi} = -\frac{f'}{2r} = R^\theta{}_{r\theta r} = R^\phi{}_{r\phi r}, \tag{A.60}$$

$$R^\phi{}_{\theta\phi\theta} = -\frac{1-f}{r^2}, \tag{A.61}$$

from which it follows that the components of the Ricci tensor are[4]

$$R_{tt} = R^r{}_{trt} + R^\theta{}_{t\theta t} + R^\phi{}_{t\phi t} = \frac{f''}{2} + \frac{f'}{2r} + \frac{f'}{2r} = 0, \tag{A.62}$$

$$R_{rr} = R^t{}_{rtr} + R^\theta{}_{r\theta r} + R^\phi{}_{r\phi r} = -\frac{f''}{2} - \frac{f'}{2r} - \frac{f'}{2r} = 0, \tag{A.63}$$

$$R_{\theta\theta} = R^t{}_{\theta t\theta} + R^r{}_{\theta r\theta} + R^\phi{}_{\theta\phi\theta} = -\frac{f'}{2r} - \frac{f'}{2r} + \frac{1-f}{r^2} = 0, \tag{A.64}$$

$$R_{\phi\phi} = R^t{}_{\phi t\phi} + R^r{}_{\phi r\phi} + R^\theta{}_{\phi\theta\phi} = -\frac{f'}{2r} - \frac{f'}{2r} + \frac{1-f}{r^2} = 0. \tag{A.65}$$

Thus, the Schwarzschild geometry is indeed a solution of the Einstein vacuum equation, as claimed.

A.4 RAIN CURVATURE

The basis 1-forms for the Schwarzschild geometry in rain coordinates are

$$\sigma^T = dT, \tag{A.66}$$

$$\sigma^R = \sqrt{\frac{2m}{r}}\, dR, \tag{A.67}$$

$$\sigma^\theta = r\, d\theta, \tag{A.68}$$

$$\sigma^\phi = r\, \sin\theta\, d\phi, \tag{A.69}$$

[4]The off-diagonal components of the Ricci tensor, such as $R^t{}_r$, are easily seen to be identically zero.

where

$$dT = dt + \frac{\sqrt{\frac{2m}{r}}}{1 - \frac{2m}{r}}\, dr, \tag{A.70}$$

$$dR = \sqrt{\frac{r}{2m}} \frac{dr}{1 - \frac{2m}{r}} + dt. \tag{A.71}$$

In these expressions, r is an implicit function of R and T, since

$$dR - dT = \left(\sqrt{\frac{r}{2m}} - \sqrt{\frac{2m}{r}}\right) \frac{dr}{1 - \frac{2m}{r}} = \sqrt{\frac{r}{2m}}\, dr \tag{A.72}$$

or equivalently

$$dr = \sqrt{\frac{2m}{r}}\,(dR - dT) = \sigma^R - \sqrt{\frac{2m}{r}}\,\sigma^T. \tag{A.73}$$

The structure equations therefore become

$$d\sigma^T = 0 = -\omega^T{}_m \wedge \sigma^m, \tag{A.74}$$

$$d\sigma^R = -\sqrt{\frac{m}{2r^3}}\, dr \wedge dR = \sqrt{\frac{m}{2r^3}}\,\sigma^T \wedge \sigma^R = -\omega^R{}_m \wedge \sigma^m, \tag{A.75}$$

$$d\sigma^\theta = dr \wedge d\theta = \left(\sigma^R - \sqrt{\frac{2m}{r}}\sigma^T\right) \wedge \frac{1}{r}\sigma^\theta = -\omega^\theta{}_m \wedge \sigma^m, \tag{A.76}$$

$$d\sigma^\phi = \sin\theta\, dr \wedge d\phi + r\, \cos\theta\, d\theta \wedge d\phi = -\omega^\phi{}_m \wedge \sigma^m. \tag{A.77}$$

These equations suggest that

$$\omega^R{}_T = \sqrt{\frac{m}{2r^3}}\,\sigma^R = \frac{m}{r^2}\, dR, \tag{A.78}$$

$$\omega^\theta{}_T = -\sqrt{\frac{2m}{r^3}}\,\sigma^\theta = -\sqrt{\frac{2m}{r}}\, d\theta, \tag{A.79}$$

$$\omega^\phi{}_T = -\sqrt{\frac{2m}{r^3}}\,\sigma^\phi = -\sqrt{\frac{2m}{r}}\,\sin\theta\, d\phi, \tag{A.80}$$

$$\omega^\theta{}_R = \frac{1}{r}\,\sigma^\theta = d\theta, \tag{A.81}$$

$$\omega^\phi{}_R = \frac{1}{r}\,\sigma^\phi = \sin\theta\, d\phi, \tag{A.82}$$

$$\omega^\phi{}_\theta = \frac{1}{r}\cot\theta\,\sigma^\phi = \cos\theta\, d\phi, \tag{A.83}$$

and it is easy to check that these educated guesses actually do satisfy the structure equations, remembering that

$$\begin{aligned} \omega^T{}_R &= \omega^R{}_T, & \omega^T{}_\theta &= \omega^\theta{}_T, & \omega^T{}_\phi &= \omega^\phi{}_T, \\ \omega^R{}_\theta &= -\omega^\theta{}_R, & \omega^R{}_\phi &= -\omega^\phi{}_R, & \omega^\theta{}_\phi &= -\omega^\phi{}_\theta. \end{aligned} \tag{A.84}$$

Since we are guaranteed a unique solution, we have found our connection 1-forms.

The curvature 2-forms are now given by

$$\begin{aligned} \Omega^T{}_R = \Omega^R{}_T &= d\omega^R{}_T + \omega^R{}_m \wedge \omega^m{}_T \\ &= -\frac{2m}{r^3}\, dr \wedge dR = \frac{2m}{r^3}\, \sigma^T \wedge \sigma^R, \end{aligned} \tag{A.85}$$

$$\begin{aligned} \Omega^T{}_\theta = \Omega^\theta{}_T &= d\omega^\theta{}_T + \omega^\theta{}_m \wedge \omega^m{}_T = d\omega^\theta{}_T + \omega^\theta{}_R \wedge \omega^R{}_T \\ &= \frac{m}{r^2}\, (dR - dT) \wedge d\theta + d\theta \wedge \frac{m}{r^2}\, dR \\ &= -\frac{m}{r^3}\, \sigma^T \wedge \sigma^\theta, \end{aligned} \tag{A.86}$$

$$\begin{aligned} \Omega^T{}_\phi = \Omega^\phi{}_T &= d\omega^\phi{}_T + \omega^\phi{}_m \wedge \omega^m{}_T = d\omega^\phi{}_T + \omega^\phi{}_R \wedge \omega^R{}_T \\ &= \frac{m}{r^2}\, (dR - dT) \wedge \sin\theta\, d\phi + \sin\theta\, d\phi \wedge \frac{m}{r^2}\, dR \\ &= -\frac{m}{r^3}\, \sigma^T \wedge \sigma^\phi, \end{aligned} \tag{A.87}$$

$$\begin{aligned} -\Omega^R{}_\theta = \Omega^\theta{}_R &= d\omega^\theta{}_R + \omega^\theta{}_m \wedge \omega^m{}_R = d\omega^\theta{}_R + \omega^\theta{}_T \wedge \omega^T{}_R \\ &= 0 - \sqrt{\frac{2m}{r^3}}\, \sigma^\theta \wedge \sqrt{\frac{m}{2r^3}}\, \sigma^R = -\frac{m}{r^3}\, \sigma^\theta \wedge \sigma^R, \end{aligned} \tag{A.88}$$

$$\begin{aligned} -\Omega^R{}_\phi = \Omega^\phi{}_R &= d\omega^\phi{}_R + \omega^\phi{}_m \wedge \omega^m{}_R = d\omega^\phi{}_R + \omega^\phi{}_T \wedge \omega^T{}_R \\ &= 0 - \sqrt{\frac{2m}{r^3}}\, \sigma^\phi \wedge \sqrt{\frac{m}{2r^3}}\, \sigma^R = -\frac{m}{r^3}\, \sigma^\phi \wedge \sigma^R, \end{aligned} \tag{A.89}$$

$$\begin{aligned} -\Omega^\theta{}_\phi = \Omega^\phi{}_\theta &= d\omega^\phi{}_\theta + \omega^\phi{}_m \wedge \omega^m{}_\theta = d\omega^\phi{}_\theta + \omega^\phi{}_R \wedge \omega^R{}_\theta + \omega^\phi{}_T \wedge \omega^T{}_\theta \\ &= -\sin\theta\, d\theta \wedge d\phi - \sin\theta\, d\phi \wedge d\theta + \frac{2m}{r} \sin\theta\, d\phi \wedge\, d\theta \\ &= \frac{2m}{r^3}\, \sigma^\phi \wedge \sigma^\theta, \end{aligned} \tag{A.90}$$

in agreement with the expressions given in Section 7.1.

A.5 COMPONENTS OF THE EINSTEIN TENSOR

The Einstein 3-forms are given (in four dimensions) by

$$\gamma^i = -\frac{1}{2}\Omega_{jk} \wedge *(\sigma^i \wedge \sigma^j \wedge \sigma^k). \tag{A.91}$$

Recalling that the Hodge dual of σ^i satisfies

$$\sigma^i \wedge *\sigma^i = g(\sigma^i, \sigma^i)\,\omega \tag{A.92}$$

(no sum on i), we can write

$$*\sigma^i = \frac{1}{3!}\epsilon^i{}_{jkl}\,\sigma^j \wedge \sigma^k \wedge \sigma^l, \tag{A.93}$$

where ϵ_{ijkl} is the alternating symbol in four dimensions, sometimes called the *Levi-Civita symbol*, which satisfies

$$\epsilon_{1234} = 1 \tag{A.94}$$

and is completely antisymmetric in its indices. The factor of 3! is needed to compensate for the implicit sum, and the raised index takes care of the factor of $g(\sigma^i, \sigma^i)$. Using the identity

$$-\epsilon_i{}^{pqr}\epsilon^i{}_{jkl}\sigma^j \wedge \sigma^k \wedge \sigma^l = 3!\,\sigma^p \wedge \sigma^q \wedge \sigma^r \tag{A.95}$$

leads to

$$-\epsilon_i{}^{pqr}*\sigma^i = \sigma^p \wedge \sigma^q \wedge \sigma^r, \tag{A.96}$$

and taking the Hodge dual of both sides now yields

$$*(\sigma^i \wedge \sigma^j \wedge \sigma^k) = -\epsilon_\ell{}^{ijk}(**\sigma^\ell) = -\epsilon_\ell{}^{ijk}\sigma^\ell \tag{A.97}$$

since $** = 1$ for 1-forms in four dimensions with signature $s = -1$. We have therefore shown that

$$\gamma^i = \frac{1}{2}\epsilon_\ell{}^{ijk}\,\Omega_{jk} \wedge \sigma^\ell. \tag{A.98}$$

Expanding the curvature 2-forms in terms of the Riemann tensor components, we have

$$\gamma^i = \frac{1}{4}\epsilon_\ell{}^{ijk}\,R_{jkpq}\,\sigma^p \wedge \sigma^q \wedge \sigma^\ell, \tag{A.99}$$

and we can now take the Hodge dual of both sides, obtaining

$$\begin{aligned} *\gamma^i &= \frac{1}{4}\epsilon_\ell{}^{ijk} R_{jkpq} *(\sigma^p \wedge \sigma^q \wedge \sigma^\ell) \\ &= -\frac{1}{4}\epsilon_\ell{}^{ijk}\epsilon_m{}^{pq\ell}\, R_{jkpq}\, \sigma^m \\ &= \frac{1}{4}\epsilon_\ell{}^{ijk}\epsilon^{\ell pq}{}_m\, R_{jkpq}\, \sigma^m. \end{aligned} \tag{A.100}$$

Using the identity

$$\begin{aligned} -\epsilon_\ell{}^{ijk}\epsilon^{\ell pq}{}_m = &\big(\delta^i{}_p\delta^j{}_q\delta^k{}_m + \delta^j{}_p\delta^k{}_q\delta^i{}_m + \delta^k{}_p\delta^i{}_q\delta^j{}_m \\ &- \delta^i{}_p\delta^k{}_q\delta^j{}_m - \delta^j{}_p\delta^i{}_q\delta^k{}_m - \delta^k{}_p\delta^j{}_q\delta^i{}_m\big), \end{aligned} \tag{A.101}$$

we have

$$\begin{aligned} *\gamma^i = &-\frac{1}{4}\big(R_{jm}{}^{ij}\,\sigma^m + R_{jk}{}^{jk}\,\sigma^i + R_{mk}{}^{ki}\,\sigma^m \\ &\qquad - R_{mk}{}^{ik}\,\sigma^m - R_{jm}{}^{ji}\,\sigma^m - R_{jk}{}^{kj}\,\sigma^i\big) \\ =&R_{pm}{}^{pi}\,\sigma^m - \frac{1}{2}R_{jk}{}^{jk}\,\sigma^i \\ =&\left(R_m{}^i - \frac{1}{2}R\,\delta_m{}^i\right)\sigma^m = G_m{}^i\sigma^m = G^i, \end{aligned} \tag{A.102}$$

and we have shown that components of the Einstein tensor, defined directly in terms of the curvature 2-forms, do indeed take their standard form, that is, that

$$*\vec{\gamma} = \vec{\boldsymbol{G}} \tag{A.103}$$

as claimed.

A.6 DIVERGENCE OF THE EINSTEIN TENSOR

We verify here that the divergence of the Einstein vector-valued 1-form $\vec{\boldsymbol{G}}$ vanishes, that is, that

$$d\gamma^i + \omega^i{}_j \wedge \gamma^j = 0, \tag{A.104}$$

where

$$\gamma^i = \frac{1}{2}\epsilon_\ell{}^{ijk}\,\Omega_{jk} \wedge \sigma^\ell \tag{A.105}$$

as shown in Section A.5. Using the Bianchi identities and the structure equations to evaluate the first term, we have

$$\begin{aligned}&2(d\gamma^i + \omega^i{}_j \wedge \gamma^j)\\ =&\epsilon_l{}^{ijk}\left(-\omega_{jm} \wedge \Omega^m{}_k \wedge \sigma^l + \Omega_{jm} \wedge \omega^m{}_k \wedge \sigma^l - \Omega_{jk} \wedge \omega^l{}_m \wedge \sigma^m\right)\\ &\quad + \epsilon_l{}^{mjk}\omega^i{}_m \wedge \Omega_{jk} \wedge \sigma^l.\end{aligned} \tag{A.106}$$

Consider first the very last term of (A.106). The antisymmetry of the Levi-Civita symbol implies that l, m, j, k must be distinct, so exactly one of them must be equal to i. The antisymmetry of the connection 1-forms now implies (in an orthonormal frame) that $i \neq m$. We have therefore shown that

$$\begin{aligned}\epsilon_l{}^{mjk}\omega^i{}_m \wedge \Omega_{jk} \wedge \sigma^l &= \epsilon_l{}^{mik}\omega^i{}_m \wedge \Omega_{ik} \wedge \sigma^l + \epsilon_l{}^{mji}\omega^i{}_m \wedge \Omega_{ji} \wedge \sigma^l\\ &\quad + \epsilon_i{}^{mjk}\omega^i{}_m \wedge \Omega_{jk} \wedge \sigma^i\\ &= 2\epsilon_l{}^{mji}\omega^i{}_m \wedge \Omega_{ji} \wedge \sigma^l + \epsilon_i{}^{mjk}\omega^i{}_m \wedge \Omega_{jk} \wedge \sigma^i\end{aligned} \tag{A.107}$$

(in an orthonormal frame), where there is no sum over i. Applying similar reasoning to the first two terms on the right-hand side of (A.106), we see that m can only take on the values i and l (in an orthonormal frame). Thus,

$$\begin{aligned}&\epsilon_l{}^{ijk}\left(-\omega_{jm} \wedge \Omega^m{}_k \wedge \sigma^l + \Omega_{jm} \wedge \omega^m{}_k \wedge \sigma^l\right)\\ &\quad = 2\epsilon_l{}^{ijk}\Omega_{jm} \wedge \omega^m{}_k \wedge \sigma^l\\ &\quad = 2\epsilon_l{}^{ijk}\Omega_{ji} \wedge \omega^i{}_k \wedge \sigma^l + 2\epsilon_l{}^{ijk}\Omega_{jl} \wedge \omega^l{}_k \wedge \sigma^l\end{aligned} \tag{A.108}$$

(in an orthonormal frame), where there is a sum over l (and of course also over j and k), but not over i. Relabeling indices and reordering factors brings this to the form

$$\begin{aligned}&\epsilon_l{}^{ijk}\left(-\omega_{jm} \wedge \Omega^m{}_k \wedge \sigma^l + \Omega_{jm} \wedge \omega^m{}_k \wedge \sigma^l\right)\\ &\quad = 2\epsilon_l{}^{ijm}\omega^i{}_m \wedge \Omega_{ji} \wedge \sigma^l + 2\epsilon_l{}^{ijk}\Omega_{jl} \wedge \omega^l{}_k \wedge \sigma^l.\end{aligned} \tag{A.109}$$

Finally, in the remaining term of (A.106), m can be i, j, or k, but not l (in an orthonormal frame), leading to

$$\begin{aligned}-\epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_m \wedge \sigma^m &= -\epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_i \wedge \sigma^i - \epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_j \wedge \sigma^j\\ &\quad - \epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_k \wedge \sigma^k\\ &= -\epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_i \wedge \sigma^i - 2\epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_j \wedge \sigma^j\\ &= -\epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_i \wedge \sigma^i - 2\epsilon_l{}^{ijk}\Omega_{jk} \wedge \omega^l{}_j \wedge \sigma^j,\end{aligned} \tag{A.110}$$

where there is a sum over j and k (and l), but not i. Relabeling indices and reordering factors leads to

$$-\epsilon_l{}^{ijk}\Omega_{jk}\wedge\omega^l{}_m\wedge\sigma^m = \epsilon_m{}^{ijk}\omega^m{}_i\wedge\Omega_{jk}\wedge\sigma^i - 2\epsilon_k{}^{ilj}\Omega_{lj}\wedge\omega^k{}_l\wedge\sigma^l. \quad (A.111)$$

If we now add up (A.107), (A.109), and (A.111), and carefully raise and lower (and reorder) some indices, everything cancels, and we have verified that (A.104) holds in an orthonormal basis. A similar computation, using the general expression of metric compatibility, establishes the same result in an arbitrary basis.

A.7 DIVERGENCE OF THE METRIC IN TWO DIMENSIONS

Consider two-dimensional Euclidean space, for which

$$d\vec{\boldsymbol{r}} = dx\,\hat{\boldsymbol{x}} + dy\,\hat{\boldsymbol{y}} \quad (A.112)$$

so that

$$*d\vec{\boldsymbol{r}} = dy\,\hat{\boldsymbol{x}} - dx\,\hat{\boldsymbol{y}} \quad (A.113)$$

and clearly

$$d{*}d\vec{\boldsymbol{r}} = \vec{\mathbf{0}}. \quad (A.114)$$

This computation shows that (A.114) must in fact hold in flat space in any signature. We check this explicitly in polar coordinates, for which

$$d\vec{\boldsymbol{r}} = dr\,\hat{\boldsymbol{r}} + r\,d\phi\,\hat{\boldsymbol{\phi}} \quad (A.115)$$

so that

$$*d\vec{\boldsymbol{r}} = r\,d\phi\,\hat{\boldsymbol{r}} - dr\,\hat{\boldsymbol{\phi}}, \quad (A.116)$$

and we compute

$$\begin{aligned} d{*}d\vec{\boldsymbol{r}} &= dr\wedge d\phi\,\hat{\boldsymbol{r}} - r\,d\phi\wedge d\hat{\boldsymbol{r}} + dr\wedge d\hat{\boldsymbol{\phi}} \\ &= dr\wedge d\phi\,\hat{\boldsymbol{r}} + r\,d\phi\wedge d\phi\,\hat{\boldsymbol{\phi}} - dr\wedge d\phi\,\hat{\boldsymbol{r}} \\ &= \vec{\mathbf{0}} \end{aligned} \quad (A.117)$$

as expected.

A similar computation shows that (A.114) holds in *any* two-dimensional geometry. We have

$$d\vec{\boldsymbol{r}} = \sigma^i\,\hat{\boldsymbol{e}}_i \quad (A.118)$$

so that

$$*d\vec{\boldsymbol{r}} = *\sigma^i\,\hat{\boldsymbol{e}}_i = \epsilon^i{}_j\,\sigma^j\,\hat{\boldsymbol{e}}_i, \tag{A.119}$$

and we compute

$$\begin{aligned} d{*}d\vec{\boldsymbol{r}} &= \epsilon^i{}_j\,d\sigma^j\,\hat{\boldsymbol{e}}_i - \epsilon^i{}_j\,\sigma^j \wedge d\hat{\boldsymbol{e}}_i \\ &= -\epsilon^i{}_j\,\omega^j{}_k \wedge \sigma^k\,\hat{\boldsymbol{e}}_i - \epsilon^i{}_j\,\sigma^j \wedge \omega^k{}_i\,\hat{\boldsymbol{e}}_k \\ &= -\sum_{ij}\left(\epsilon^i{}_j\,\omega^j{}_i \wedge \sigma^i\,\hat{\boldsymbol{e}}_i + \epsilon^i{}_j\,\sigma^j \wedge \omega^j{}_i\,\hat{\boldsymbol{e}}_j\right) \\ &= -\sum_{ij}\left(\epsilon^i{}_j\,\omega^j{}_i \wedge \sigma^i\,\hat{\boldsymbol{e}}_i + \epsilon^j{}_i\,\sigma^i \wedge \omega^i{}_j\,\hat{\boldsymbol{e}}_i\right) \\ &= -\sum_{ij}\left(\epsilon^i{}_j\,\omega^j{}_i \wedge \sigma^i\,\hat{\boldsymbol{e}}_i + \epsilon^i{}_j\,\sigma^i \wedge \omega^j{}_i\,\hat{\boldsymbol{e}}_i\right) \\ &= \vec{\boldsymbol{0}}, \end{aligned} \tag{A.120}$$

where we have used the fact that

$$\epsilon^i{}_j\,\omega^j{}_i = \epsilon^j{}_i\,\omega^i{}_j \qquad \text{(no sum)} \tag{A.121}$$

for *each* value of i and j.

A similar computation can be done in any dimension; the divergence of the metric always vanishes, that is, (A.114) is always true.

A.8 DIVERGENCE OF THE METRIC

We now consider the general case, for which

$$d\vec{\boldsymbol{r}} = \sigma^i\,\hat{\boldsymbol{e}}_i. \tag{A.122}$$

Recalling that the Hodge dual of σ^i satisfies

$$\sigma^i \wedge *\sigma^i = g(\sigma^i, \sigma^i)\,\omega, \tag{A.123}$$

we can write

$$*\sigma^i = \frac{1}{(n-1)!}\,\epsilon^i{}_{j\ldots k}\,\sigma^j \wedge \ldots \wedge \sigma^k, \tag{A.124}$$

where the n-index object $\epsilon_{ij\ldots k}$ is the alternating symbol in n dimensions, sometimes called the *Levi-Civita symbol*, which satisfies

$$\epsilon_{1\ldots n} = 1 \tag{A.125}$$

and is completely antisymmetric in its indices. The factor of $(n-1)!$ is needed to compensate for the implicit sum, and the raised index takes care of the factor of $g(\sigma^i, \sigma^i)$. We therefore have

$$*d\vec{\boldsymbol{r}} = \frac{1}{(n-1)!}\, \epsilon^i{}_{j\ldots k}\, \hat{\boldsymbol{e}}_i\, \sigma^j \wedge \ldots \wedge \sigma^k \tag{A.126}$$

so that, using the structure equations,

$$\begin{aligned} d{*}d\vec{\boldsymbol{r}} &= \frac{1}{(n-1)!}\, \epsilon^i{}_{j\ldots k} \Big(d\hat{\boldsymbol{e}}_i \wedge \sigma^j \wedge \ldots \wedge \sigma^k + \hat{\boldsymbol{e}}_i\, d\sigma^j \wedge \ldots \wedge \sigma^k + \ldots \\ &\qquad\qquad + (-1)^{n-2}\, \hat{\boldsymbol{e}}_i\, \sigma^j \wedge \ldots \wedge d\sigma^k \Big) \\ &= \frac{1}{(n-1)!}\, \epsilon^i{}_{j\ldots k} \Big(\hat{\boldsymbol{e}}_m\, \omega^m{}_i \wedge \sigma^j \wedge \ldots \wedge \sigma^k - \hat{\boldsymbol{e}}_i\, \omega^j{}_m \wedge \sigma^m \wedge \ldots \wedge \sigma^k - \ldots \\ &\qquad\qquad - (-1)^{n-2}\, \hat{\boldsymbol{e}}_i\, \sigma^j \wedge \ldots \wedge \omega^k{}_m \wedge \sigma^m \Big) \\ &= \frac{1}{(n-1)!}\, \hat{\boldsymbol{e}}_i \Big(\epsilon^m{}_{j\ldots k} \omega^i{}_m - \epsilon^i{}_{m\ldots k} \omega^m{}_j - \ldots \\ &\qquad\qquad - \epsilon^i{}_{j\ldots m} \omega^m{}_k \Big) \wedge \sigma^j \wedge \ldots \wedge \sigma^k \end{aligned} \tag{A.127}$$

since moving the connection 1-forms to the front eliminates the signs, and the indices being summed over can be relabeled. To see that this expression is zero, consider the factor in parentheses. Lower an index to obtain

$$\begin{aligned} \epsilon^m{}_{j\ldots k} \omega_{im} - \ \epsilon_{im\ldots k} \omega^m{}_j - \epsilon_{ij\ldots m} \omega^m{}_k & \\ = \epsilon^m{}_{j\ldots k} \omega_{im} + \epsilon^m{}_{i\ldots k} \omega_{mj} + \epsilon^m{}_{j\ldots i} \omega_{mk}. & \end{aligned} \tag{A.128}$$

If $i,j,\ldots,k$ are distinct, then each term vanishes for each value of m. For the first term to be nonzero, we must have $i \in \{j \ldots k\}$, and since i is a free index, this must be true in the remaining terms as well. But for each allowed choice of i, precisely one of the remaining terms is nonzero, and the antisymmetry of the connection 1-forms (due to metric compatibility) now ensures that this nonzero term exactly cancels the first term.

This completes the proof that $d{*}d\vec{\boldsymbol{r}} = 0$; the metric is divergence free.

A.9 ROBERTSON–WALKER CURVATURE

The Robertson–Walker line element is

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2\theta\, d\phi^2 \right) \right). \tag{A.129}$$

It now follows immediately that

$$d\vec{\boldsymbol{r}} = dt\,\hat{\boldsymbol{t}} + \frac{a(t)\,dr}{\sqrt{1-\kappa r^2}}\,\hat{\boldsymbol{r}} + a(t)\,r\,d\theta\,\hat{\boldsymbol{\theta}} + a(t)\,r\,\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}} \tag{A.130}$$

and the basis 1-forms are

$$\sigma^t = dt, \tag{A.131}$$

$$\sigma^r = \frac{a(t)\,dr}{\sqrt{1-\kappa r^2}}, \tag{A.132}$$

$$\sigma^\theta = a(t)\,r\,d\theta, \tag{A.133}$$

$$\sigma^\phi = a(t)\,r\,\sin\theta\,d\phi. \tag{A.134}$$

The structure equations therefore become

$$d\sigma^t = 0 = -\omega^t{}_m \wedge \sigma^m, \tag{A.135}$$

$$d\sigma^r = \frac{\dot{a}\,dt\wedge dr}{\sqrt{1-\kappa r^2}} = -\omega^r{}_m \wedge \sigma^m, \tag{A.136}$$

$$d\sigma^\theta = \dot{a}r\,dt\wedge d\theta + a\,dr\wedge d\theta = -\omega^\theta{}_m \wedge \sigma^m, \tag{A.137}$$

$$\begin{aligned} d\sigma^\phi &= \dot{a}r\,\sin\theta\,dt\wedge d\phi + a\,\sin\theta\,dr\wedge d\phi + ar\,\cos\theta\,d\theta\wedge d\phi \\ &= -\omega^\phi{}_m \wedge \sigma^m. \end{aligned} \tag{A.138}$$

These equations suggest that

$$\omega^r{}_t = \frac{\dot{a}\,dr}{\sqrt{1-\kappa r^2}} = \frac{\dot{a}}{a}\,\sigma^r, \tag{A.139}$$

$$\omega^\theta{}_t = r\dot{a}\,d\theta = \frac{\dot{a}}{a}\,\sigma^\theta, \tag{A.140}$$

$$\omega^\theta{}_r = \sqrt{1-\kappa r^2}\,d\theta = \frac{\sqrt{1-\kappa r^2}}{ar}\,\sigma^\theta, \tag{A.141}$$

$$\omega^\phi{}_t = r\dot{a}\,\sin\theta\,d\phi = \frac{\dot{a}}{a}\,\sigma^\phi, \tag{A.142}$$

$$\omega^\phi{}_r = \sqrt{1-\kappa r^2}\,\sin\theta\,d\phi = \frac{\sqrt{1-\kappa r^2}}{ar}\,\sigma^\phi, \tag{A.143}$$

$$\omega^\phi{}_\theta = \cos\theta\,d\phi = \frac{\cot\theta}{ar}\,\sigma^\phi, \tag{A.144}$$

and it is easy to check that these educated guesses actually do satisfy the structure equations, remembering that

$$\begin{aligned} &\omega^r{}_t = \omega^t{}_r, && \omega^\theta{}_t = \omega^t{}_\theta, && \omega^\phi{}_t = \omega^t{}_\phi, \\ &\omega^r{}_\theta = -\omega^\theta{}_r, && \omega^r{}_\phi = -\omega^\phi{}_r, && \omega^\theta{}_\phi = -\omega^\phi{}_\theta. \end{aligned} \tag{A.145}$$

Since we are guaranteed a unique solution, we have found our connection 1-forms.

The curvature 2-forms are now given by

$$\Omega^t{}_r = d\omega^t{}_r + \omega^t{}_m \wedge \omega^m{}_r = \frac{\ddot{a}\,dt \wedge dr}{\sqrt{1-\kappa r^2}} = \frac{\ddot{a}}{a}\sigma^t \wedge \sigma^r, \tag{A.146}$$

$$\Omega^t{}_\theta = d\omega^t{}_\theta + \omega^t{}_m \wedge \omega^m{}_\theta = \ddot{a}r\,dt \wedge d\theta = \frac{\ddot{a}}{a}\sigma^t \wedge \sigma^\theta, \tag{A.147}$$

$$\Omega^t{}_\phi = d\omega^t{}_\phi + \omega^t{}_m \wedge \omega^m{}_\phi = \ddot{a}r\,\sin\theta\,dt \wedge d\phi = \frac{\ddot{a}}{a}\sigma^t \wedge \sigma^\phi, \tag{A.148}$$

$$\Omega^r{}_\theta = d\omega^r{}_\theta + \omega^r{}_m \wedge \omega^m{}_\theta = \frac{(\dot{a}^2+\kappa)r\,dr \wedge d\theta}{\sqrt{1-\kappa r^2}} = \frac{\dot{a}^2+\kappa}{a^2}\sigma^r \wedge \sigma^\theta, \tag{A.149}$$

$$\begin{aligned}\Omega^r{}_\phi &= d\omega^r{}_\phi + \omega^r{}_m \wedge \omega^m{}_\phi = \frac{(\dot{a}^2+\kappa)r\,\sin\theta\,dr \wedge d\phi}{\sqrt{1-\kappa r^2}} \\ &= \frac{\dot{a}^2+\kappa}{a^2}\sigma^r \wedge \sigma^\phi,\end{aligned} \tag{A.150}$$

$$\begin{aligned}\Omega^\theta{}_\phi &= d\omega^\phi{}_\theta + \omega^\phi{}_m \wedge \omega^m{}_\theta = (\dot{a}^2+\kappa)r^2\,\sin\theta\,d\theta \wedge d\phi \\ &= \frac{\dot{a}^2+\kappa}{a^2}\sigma^\theta \wedge \sigma^\phi.\end{aligned} \tag{A.151}$$

Again we have

$$\begin{array}{lll}\Omega^r{}_t = \Omega^t{}_r, & \Omega^\theta{}_t = \Omega^t{}_\theta, & \Omega^\phi{}_t = \Omega^t{}_\phi \\ \Omega^r{}_\theta = -\Omega^\theta{}_r, & \Omega^r{}_\phi = -\Omega^\phi{}_r, & \Omega^\theta{}_\phi = -\Omega^\phi{}_\theta.\end{array} \tag{A.152}$$

A.10 Birkhoff's Theorem

Birkhoff's Theorem states that the Schwarzschild geometry is the *only* spherically symmetric solution of Einstein's vacuum equation. This result is remarkable, in that the Schwarzschild geometry has a timelike symmetry (Killing vector), even though this was not assumed; spherically symmetric vacuum solutions of Einstein's equation are automatically time independent! We outline the proof of Birkhoff's Theorem below.

Step 1: Spherical symmetry implies that our spacetime can be foliated by 2-spheres, each with line element

$$ds^2 = r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right). \tag{A.153}$$

We will use r as our third coordinate and call the remaining coordinate t. Spherical symmetry further implies that we can choose the r and t direc-

tions to be orthogonal to our spheres. The full line element therefore takes the form[5]

$$ds^2 = -A^2\,dt^2 + B^2 dr^2 + 2C\,dt\,dr + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right), \qquad \text{(A.154)}$$

where A, B, and C are functions of t and r only.

Step 2: We can write the metric in the (t,r)-plane as the difference of squares. With θ and ϕ constant we have

$$\begin{aligned} ds^2 &= -A^2\,dt^2 + B^2 dr^2 + 2C\,dt\,dr \\ &= -\left(A\,dt - \frac{C}{A}\,dr\right)^2 + \left(B^2 - \frac{C^2}{A^2}\right) dr^2. \end{aligned} \qquad \text{(A.155)}$$

But any differential in two dimensions is a multiple of an exact differential, so we must have[6]

$$A\,dt - \frac{C}{A}\,dr = P\,dt' \qquad \text{(A.156)}$$

for some P and t'. We can therefore write the line element as

$$ds^2 = -P^2\,dt'^2 + Q^2\,dr^2, \qquad \text{(A.157)}$$

where $Q^2 = B^2 - \frac{C^2}{A^2}$ (which is positive since the signature is invariant), and where P and Q are now to be thought of as functions of t' and r. We henceforth drop the primes; this argument shows that we could have simply assumed that $C = 0$ (in which case $P = A$ and $Q = B$).

Step 3: Compute the Einstein tensor for the line element

$$ds^2 = -A^2\,dt^2 + B^2\,dr^2 + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right), \qquad \text{(A.158)}$$

[5]By choosing A^2 and B^2 as coefficients, we are making certain assumptions about signature, which turn out to imply that $r > 2m$. These assumptions do not affect the argument in any way; a more general derivation would simply rename these coefficients as A and B.

[6]One way to see this is to use existence and uniqueness properties of differential equations. Assuming that the differential equation $\frac{dt}{dr} = \frac{C}{A^2}$ has a solution $t = T(r)$, then

$$d(t - T(r)) = dt - \frac{C}{A^2}\,dr = \frac{A\,dt - \frac{C}{A}\,dr}{A}$$

and we can set $t' = t - T(r)$ and $P = A$.

where A and B are functions of t and r. The nonzero components are

$$G^t{}_t = -\frac{1}{r^2 B^2}\left(B^2 - 1 + \frac{2r}{B}\frac{\partial B}{\partial r}\right), \tag{A.159}$$

$$G^t{}_r = -\frac{2}{r\,A^2 B}\frac{\partial B}{\partial t}, \tag{A.160}$$

$$G^r{}_r = -\frac{1}{r^2 B^2}\left(B^2 - 1 - \frac{2r}{A}\frac{\partial A}{\partial r}\right), \tag{A.161}$$

$$\begin{aligned} G^\theta{}_\theta = G^\phi{}_\phi = \frac{1}{r^2 B^2}\Bigg(&-\frac{r}{B}\frac{\partial B}{\partial r} + \frac{r}{A}\frac{\partial A}{\partial r} + \frac{r^2}{A}\frac{\partial^2 A}{\partial r^2} - \frac{r^2}{B}\frac{\partial^2 B}{\partial t^2} \\ &-\frac{r^2}{AB}\frac{\partial A}{\partial r}\frac{\partial B}{\partial r} + \frac{r^2}{AB}\frac{\partial A}{\partial t}\frac{\partial B}{\partial t}\Bigg). \end{aligned} \tag{A.162}$$

Step 4: Setting $G^t{}_r = 0$, we see that B must be a function of r alone.

Step 5: Setting $G^t{}_t - G^r{}_r = 0$, we obtain

$$\frac{1}{B}\frac{\partial B}{\partial r} = -\frac{1}{A}\frac{\partial A}{\partial r}, \tag{A.163}$$

from which it follows that

$$B = \frac{\lambda}{A}, \tag{A.164}$$

where λ is an integration "constant," which, however, might be a function of t. However, by setting $dt' = \frac{dt}{\lambda}$, we can absorb this integration constant into the definition of t. We can therefore safely assume that $\lambda = 1$, so that

$$B = \frac{1}{A} \tag{A.165}$$

and the line element takes the form

$$ds^2 = -A^2\,dt^2 + \frac{1}{A^2}\,dr^2 + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right), \tag{A.166}$$

where A is a function of r alone.

Step 6: Setting $G^t{}_t = 0$ and using (A.163) and (A.165) we obtain

$$\frac{1}{A}\frac{dA}{dr} = -\frac{1}{B}\frac{dB}{dr} = \frac{B^2-1}{2r} = \frac{1-A^2}{2r\,A^2}, \tag{A.167}$$

which can be separated to yield

$$\frac{A\,dA}{1-A^2} = \frac{dr}{2r}. \tag{A.168}$$

Integrating both sides now leads to

$$-\ln(1 - A^2) = \ln(ar), \tag{A.169}$$

where a is an integration constant, or equivalently

$$A^2 = 1 - \frac{1}{ar}. \tag{A.170}$$

The integration constant can be related to mass by comparing the Newtonian limit of this geometry to the (Newtonian) gravitational field of a point particle, leading to

$$\frac{1}{a} = 2m. \tag{A.171}$$

Inserting (A.165), (A.170), and (A.171) into (A.166), we recover the Schwarzschild line element in its usual form.

Step 7: Finally, inserting $G^t{}_t = 0$ into $G^\theta{}_\theta = G^\phi{}_\phi$ and using (A.163) along with the relationship

$$\frac{\partial}{\partial r}\left(\frac{1}{A}\frac{\partial A}{\partial r}\right) = \frac{1}{A}\frac{\partial^2 A}{\partial r^2} - \frac{1}{A^2}\left(\frac{\partial A}{\partial r}\right)^2 \tag{A.172}$$

shows that these components vanish identically, thus completing the proof.

A.11 THE STRESS TENSOR FOR A POINT CHARGE

As discussed in Section 15.9, the electromagnetic field is a 2-form F. Maxwell's equations imply that F is closed, that is, that

$$dF = 0, \tag{A.173}$$

which leads to the assumption that

$$F = dA, \tag{A.174}$$

where A is the *4-potential.* In special relativity, we have

$$A = -\Phi\, dt + \vec{\boldsymbol{A}} \cdot d\vec{\boldsymbol{r}}, \tag{A.175}$$

where Φ is the *electric potential*, also called the *scalar potential*, and the 3-vector $\vec{\boldsymbol{A}}$ is the *magnetic potential*, also called the *vector potential.* The electric and magnetic fields are then

$$\vec{\boldsymbol{E}} = -\vec{\boldsymbol{\nabla}}\Phi, \tag{A.176}$$

$$\vec{\boldsymbol{B}} = \vec{\boldsymbol{\nabla}} \times \vec{\boldsymbol{A}}, \tag{A.177}$$

both of which are 3-vectors, that is, have no component in the t direction. In the language of differential forms, we have (compare to (15.69))

$$F = E \wedge dt + *B, \tag{A.178}$$

where of course $E = \vec{\boldsymbol{E}} \cdot d\vec{\boldsymbol{r}}$, $B = \vec{\boldsymbol{B}} \cdot d\vec{\boldsymbol{r}}$, and where $*$ denotes the *three-dimensional* Hodge dual operator.

The electric field of a point charge should take the form

$$\vec{\boldsymbol{E}} = \frac{q}{r^2}\,\hat{\boldsymbol{r}}, \tag{A.179}$$

where in our geometric units we set $4\pi\epsilon_0 = 1$, so that, like mass, charge has the dimensions of length. We therefore expect the electromagnetic field of a point charge to take the form[7]

$$F = \frac{1}{2} F_{ij}\,\sigma^i \wedge \sigma^j = \frac{q}{r^2}\,\sigma^r \wedge \sigma^t. \tag{A.180}$$

The components of the stress-energy tensor for an electromagnetic field can be shown to be

$$4\pi T^i{}_j = F^{im} F_{jm} - \frac{1}{4}\delta^i{}_j F^{mn} F_{mn}, \tag{A.181}$$

which reduces to the diagonal matrix

$$4\pi \left(T^i{}_j\right) = \frac{1}{2}\begin{pmatrix} -q/r^2 & 0 & 0 & 0 \\ 0 & -q/r^2 & 0 & 0 \\ 0 & 0 & q/r^2 & 0 \\ 0 & 0 & 0 & q/r^2 \end{pmatrix} \tag{A.182}$$

for the electromagnetic field of a point charge.

[7] If this expression reminds you of curvature 2-forms, that is no coincidence. *Gauge field theory*, including electromagnetism, can be formulated in the geometric language of *fiber bundles*; the corresponding notion of curvature in the electromagnetic case is precisely F.

≻ Appendix B ≺

Tensor Notation

For the benefit of readers who wish to compare the results presented in this book using differential forms with similar results elsewhere expressed in terms of tensor language, we outline here a rough dictionary between several different notations.

B.1 Invariant Language

Mathematicians tend to use a basis-independent language to describe tensor analysis.

Vector fields $\vec{\boldsymbol{v}}$, often casually referred to simply as *vectors*, are defined as directional derivatives, that is,

$$\vec{\boldsymbol{v}}(f) = \vec{\boldsymbol{v}} \cdot \vec{\boldsymbol{\nabla}} f, \tag{B.1}$$

although we will follow standard practice and drop the arrows in the remainder of this section. Another basic object is a *1-form* α, defined as a linear map on vector fields, so that $\alpha(v) \in \mathbb{R}$. A special case of a 1-form is the differential df of a function f, defined via

$$df(v) = v(f). \tag{B.2}$$

Not only can 1-forms act on vectors, but vectors can act on 1-forms, via

$$v(\alpha) = \alpha(v), \tag{B.3}$$

which should not be confused with the action of v on functions, even though a similar notation is used. A *tensor* is then any multilinear map on vectors and 1-forms. More precisely, a $\binom{p}{q}$-tensor T is a multilinear map taking p 1-forms and q vectors as inputs and producing a real number at each point. T is said to be *covariant* if $p = 0$, and *contravariant* if $q = 0$.

The *tensor product* of two 1-forms α, β is a rank 2 covariant tensor $\alpha \otimes \beta$, defined by

$$(\alpha \otimes \beta)(u, v) = \alpha(u)\,\beta(v), \tag{B.4}$$

although the parentheses around $\alpha \otimes \beta$ are often omitted. The tensor product $T \otimes U$ of a $\binom{p}{q}$-tensor T with an $\binom{r}{s}$-tensor U can be defined similarly, and is a $\binom{p+r}{q+s}$-tensor.

In this language, a *differential form* is a covariant tensor that is completely antisymmetric in its arguments, so that interchanging any two of the vectors it acts on results in an overall minus sign.

B.2 COMPONENTS

Physicists tend to use components to describe tensor analysis.

Given a vector basis $\{e_1, ..., e_n\}$, the dual basis $\{\sigma^1, ..., \sigma^n\}$ of 1-forms is defined by

$$\sigma^i(e_j) = \delta^i{}_j, \tag{B.5}$$

where $\delta^i{}_j$ denotes the Kronecker delta.[1]

A vector field v can be expanded in terms of the basis as

$$v = v^i e_i \tag{B.6}$$

and the components v^i of v can be computed as

$$v^i = \sigma^i(v). \tag{B.7}$$

Similarly, a 1-form α can be expanded as

$$\alpha = \alpha_i \sigma^i \tag{B.8}$$

with

$$\alpha_i = e_i(\alpha). \tag{B.9}$$

A tensor T can also be expanded in this form, resulting in

$$T = T^{i...j}{}_{k...l}\, e_i \otimes ... \otimes e_j \otimes \sigma^k \otimes ... \otimes \sigma^l, \tag{B.10}$$

where the components of T are given by

$$T^{i...j}{}_{k...l} = T(\sigma^i, ..., \sigma^j, e_k, ..., e_l). \tag{B.11}$$

Expressions such as (B.10) are so unwieldy that one usually omits the basis elements from consideration and works just with the components.

[1]Physicists often work in an orthonormal basis, but nothing in the component formalism requires this; (B.5) is *not* a statement about orthonormality.

Thus, one often refers to $R^i{}_{jkl}$ as though it were the Riemann tensor itself, rather than the components of the Riemann tensor (or, equivalently, of the curvature 2-forms) in a particular basis.

Dropping the basis elements does, however, complicate the notion of differentiation of a tensor. From (B.10), it is clear that differentiation involves extensive use of the product rule, not merely the partial derivatives of the components. Furthermore, these product rule terms all involve derivatives of basis elements, and hence the connection. In component language, these product rule terms are treated as "correction terms" involving Christoffel symbols $\Gamma^i{}_{jk}$, the components of the connection 1-forms $\omega^i{}_j$.[2] Further discussion of the *covariant differentiation* of tensors is, however, beyond the scope of this book.

The distinction between tensors and tensor components is further blurred by the *abstract index* approach pioneered by Roger Penrose, in which one writes component expressions but does not specify any particular basis.

In component language, antisymmetrization is often denoted by square brackets around certain indices, so that a 2-form might be written as $T_{[ij]}$, and a wedge product of 1-forms might be written as $\alpha_{[i}\beta_{j]}$.

B.3 INDEX GYMNASTICS

General relativity is described by Lorentzian geometry, in which there is a metric tensor g, whose components are related to the line element

$$ds^2 = g_{ij}\, dx^i\, dx^j \tag{B.12}$$

via

$$g = g_{ij}\, dx^i \otimes dx^j. \tag{B.13}$$

The metric allows us to identify physically equivalent tensors of different types (but the same rank) by "raising and lowering indices." First of all, the metric has an inverse, written with superscripts, that is

$$g^{ij} g_{jk} = \delta^i{}_k. \tag{B.14}$$

The components v_i of the 1-form physically equivalent to v are given by

$$v_i = g_{ij} v^j, \tag{B.15}$$

[2]The Christoffel symbols are *not* the components of some 3-index tensor. The easiest way to see this is to note that they vanish in rectangular coordinates, but not in polar coordinates, whereas a tensor that vanishes in one basis must vanish in all bases.

and similarly g^{ij} is used to "raise the index" on the components α_i of a 1-form α; (B.14) ensures that these operations are compatible.

Tensor equations must balance, that is, the same indices must appear on both sides, in the same place (up or down), apart from any *dummy indices* that are being summed over. Such repeated indices should always occur in pairs, with one up and the other down, although it makes no difference which is which.

As an example of index gymnastics, we reproduce the transition from (2.11) and (2.12) to (2.13), showing that Killing vectors lead to constants of geodesic motion. (These equations involve covariant derivatives, written with a semicolon; the only relevant properties are that derivatives involve a product rule and that the result is tensorial, that is, we can ignore the semicolon when doing our index gymnastics.)

In component notation, the geodesic equation (2.11) takes the form

$$v^i v^j{}_{;i} = 0 \tag{B.16}$$

and Killing's equation (2.12) becomes

$$X_{i;j} + X_{j;i} = 0. \tag{B.17}$$

To show that $\vec{\boldsymbol{X}} \cdot \vec{\boldsymbol{v}}$ is constant along the geodesic, we note first of all that the dot product is represented as

$$g(X, v) = g_{ij}\, X^i\, v^j = X_j\, v^j. \tag{B.18}$$

Taking a directional derivative along v, we obtain

$$\begin{aligned} v^k \left(X_j\, v^j\right)_{;k} &= v^k \left(X_{j;k}\, v^j + X_j\, v^j{}_{;k}\right) \\ &= X_{j;k}\, v^j\, v^k + X_j\, v^k\, v^j{}_{;k} \\ &= \frac{1}{2}\left(X_{j;k}\, v^j\, v^k + X_{k;j}\, v^k\, v^j\right) + X_j\, v^k\, v^j{}_{;k} \\ &= \frac{1}{2}\left(X_{j;k} + X_{k;j}\right) v^j\, v^k + X_j\, v^k\, v^j{}_{;k} \\ &= 0 + 0, \end{aligned} \tag{B.19}$$

where we have used the product rule in the first equality, and the freedom to rename dummy indices in the third equality.

Annotated Bibliography

General Relativity

- Edwin F. Taylor and John Archibald Wheeler, **Exploring Black Holes**, Addison Wesley Longman, Boston, 2000. *An elementary introduction to the relativity of black holes, using line elements. Not easy to read, but worth it.*
- James B. Hartle, **Gravity**, Addison Wesley, Boston, 2003. *An "examples first" introduction to general relativity, discussing applications of Einstein's equations before presenting the mathematics behind the equations.*
- Ray d'Inverno, **Introducing Einstein's Relativity**, Oxford University Press, Oxford, 1991. *An excellent introduction to general relativity, which also covers some topics not usually seen in introductory texts.*
- David McMahon, **Relativity Demystified**, McGraw Hill, New York, 2006. *An abridged treatment of relativity, containing a remarkably complete collection of formulas and topics, without much derivation. A useful reference.*
- Sean Carroll, **Spacetime and Geometry: An Introduction to General Relativity**, Addison Wesley, Boston, 2004. *An excellent traditional but somewhat sophisticated introduction to general relativity.*
- Bernard F. Schutz, **A First Course in General Relativity**, Cambridge University Press, Cambridge, 1985. *A good, easy introduction to the basics of both tensors and general relativity.*
- Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, **Gravitation**, Freeman, San Fransisco, 1973. *The physicist's bible of general relativity. Exhaustively complete.*
- Robert M. Wald, **General Relativity**, University of Chicago Press, Chicago, 1984. *The best available introduction to general relativity for advanced students, but uses sophisticated notation (which has become the standard for researchers in the field).*
- Rainer K. Sachs, **General Relativity for Mathematicians**, Springer, New York, 1977. *A very pure mathematical treatment of general relativity. Requires a strong background in differential geometry.*

DIFFERENTIAL FORMS

- Harley Flanders, **Differential Forms with Applications to the Physical Sciences**, Academic Press, New York, 1963; Dover, New York, 1989. *Excellent but somewhat dated introduction to differential forms; only covers the positive definite case.*

- Richard L. Bishop and Samuel I. Goldberg, **Tensor Analysis on Manifolds**, Macmillan, New York, 1968; Dover, New York, 1980. *Somewhat more sophisticated; a good reference for tensors.*

- Robert H. Wasserman, **Tensors and Manifolds**, Oxford University Press, Oxford, 1992 (2nd edition: 2009). *A more readable version of Bishop and Goldberg; but not as cheap.*

- David Bachman, **A Geometric Approach to Differential Forms**, Birkhäuser, Boston, 2006. *Very readable math text.*

- David Lovelock and Hanno Rund, **Tensors, Differential Forms, and Variational Principles**, Wiley, New York, 1975; Dover, New York, 1989. *Another inexpensive Dover reprint, covering similar material from a slightly different point of view.*

- Hwei Hsu, **Vector Analysis** (Chapter 10 only), Simon and Schuster Technical Outlines, New York, 1969. *A sort-of Schaum outline for differential forms.*

- H. M. Schey, **Div, Grad, Curl, and All That**, Norton, New York, 1973 (2nd edition: 1992; 3rd edition: 1997; 4th edition: 2005) *Excellent informal introduction to the geometry of vector calculus.*

- Gabriel Weinreich, **Geometrical Vectors**, University of Chicago Press, Chicago, 1998. *A nice (if idiosyncratic), geometrical description of differential forms without ever using those words.*

- Barrett O'Neill, **Elementary Differential Geometry**, Academic Press, New York, 1966 (2nd edition: 1997). *Standard, fairly readable introduction to differential geometry in ordinary Euclidean 3-space. Uses differential forms extensively.*

- Richard S. Millman and George D. Parker, **Elements of Differential Geometry**, Prentice-Hall, Englewood Cliffs, NJ, 1977 (2nd edition: 1997). *Good undergraduate text covering similar material to the above. Does not use differential forms.*

- Barrett O'Neill, **Semi-Riemannian Geometry**, Academic Press, New York, 1983. *Graduate math text. The best available treatment of differential geometry without the usual assumption that the metric is positive definite.*

- William M. Boothby, **An Introduction to Differentiable Manifolds and Riemannian Geometry**, Academic Press, New York, 1986. *Fairly readable graduate math text. Good but brief treatment of differential forms and integration, but emphasis is on Lie groups.*

- Manfredo Perdigão do Carmo, **Riemannian Geometry**, Birkhäuser, Boston, 1992. *Quite readable math text, but no discussion of differential forms.*
- Michael Spivak, **A Comprehensive Introduction to Differential Geometry** (5 volumes), Publish or Perish, Houston, 1970-1975 (2nd edition: 1979). *Everything you ever wanted to know, and then some, but not easy to read.*
- Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, **Gravitation**, Freeman, San Fransisco, 1973. *Don't underestimate this classic! Chapter 4 has a nice introduction to differential forms, including great pictures and a discussion of electromagnetism. Part III also discusses the differential geometry needed for relativity.*
- Ian M. Benn and Robin W. Tucker, **An Introduction to Spinors and Geometry, with Applications to Physics**, Adam Hilger, Bristol, 1988. *A very pure mathematical treatment of spinors in the language of differential forms.*

References

[1] Tevian Dray and Corinne A. Manogue, **The Geometry of Vector Calculus**, http://www.math.oregonstate.edu/BridgeBook, 2009–2013.

[2] Tevian Dray, **The Geometry of Special Relativity**, A K Peters/CRC Press, Boca Raton, FL, 2012.

[3] Karl Schwarzschild, "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie," *Sitzber. Deut. Akad. Wiss. Berlin* **7**, 189–196, 1916.

[4] Johannes Droste, "Het zwaartekrachtsveld van een of meer lichamen volgens de theorie van Einstein," Proefschrift (doctoral dissertation), Rijksuniversiteit te Leiden, 1916.

[5] Edwin F. Taylor and John Archibald Wheeler, **Exploring Black Holes**, Addison Wesley Longman, Boston, 2000.

[6] Tevian Dray, "The Asymptotic Structure of a Family of Einstein-Maxwell Solutions," Ph.D. dissertation, University of California at Berkeley, 1981.

[7] Abhay Ashtekar and Tevian Dray, "On the Existence of Solutions to Einstein's Equation with Non-Zero Bondi News," *Commun. Math. Phys.* **79**, 581–589 (1981).

[8] Tevian Dray, "On the Asymptotic Flatness of the C-Metrics at Spatial Infinity," *Gen. Rel. Grav.* **14**, 109–112 (1982).

[9] C Cutler and R M Wald, "Existence of radiating Einstein–Maxwell Solutions which are C^∞ on all of I^+ and I^0," *Class. Quantum Grav.* **6**, 453–466 (1989).

[10] Ray d'Inverno, **Introducing Einstein's Relativity**, Oxford University Press, Oxford, 1991.

[11] Stephen W. Hawking and George F. R. Ellis, **The Large Scale Structure of Space-Time**, Cambridge University Press, Cambridge, 1973.

[12] Gabriel Weinreich, **Geometrical Vectors**, University of Chicago Press, Chicago, 1998.

[13] David Bachman, **A Geometric Approach to Differential Forms**, Birkhäuser, Boston, 2006.

[14] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler, **Gravitation**, Freeman, San Fransisco, 1973.

[15] Jürgen Ehlers, "Christoffel's Work on the Equivalence Problem for Riemannian Spaces and Its Importance for Modern Field Theories of Physics," in: **E. B. Christoffel**, ed. P. L. Butzer and F. Fehér, Birkhäuser Verlag, Basel, 1981.

[16] Anders Karlhede, "A Review of the Geometrical Equivalence of Metrics in General Relativity," *Gen. Rel. Grav.* **12**, 693–707 (1980).

[17] Ian M. Benn and Robin W. Tucker, **An Introduction to Spinors and Geometry, with Applications to Physics**, Adam Hilger, Bristol, 1988.

Index